Universitat Politècnica de Catalunya

Treball Final de Grau

Bachelor Degree in Informatics Engineering

Speech2Anim

Animació procedural d'avatars basada en anàlisi de vídeos

Autor:
Gerard del Castillo Lite

Director:
Carlos Andújar Gran

Codirector:
Nuria Pelechano

Q1 Curs 2017-2018

Abstract

English

In this document we describe a platform to aid in the production of one of the most time consuming aspects of computer generated images: animation. This platform uses state-of-the-art tools from multiple fields such as *Machine Learning* and *Computer Vision* to provide a framework in which users can explore the correlation between certain movements and speech features on real training videos. This will aid in the task to apply prediction models to create or modify existing animations according to an input audio. We focus on creating a modular and extensible tool for automatizing the training and animation processes rather than actual development of final prediction models, which require extensive data sets, large computing resources, and a deep knowledge of character animation. With this project, we succeeded in creating a modular and extensible tool which automatizes the training and animation process as a *Blender* add-on.

Català

En aquest document descrivim una plataforma per ajudar en la producció d'un dels aspectes que consumeixen més temps de la generació d'imatges per ordinador: l'animació. Aquesta plataforma utilitza eines d'avantguarda de múltiples camps com *Machine Learning* i *Computer Vision* per proporcionar un marc en el qual els usuaris puguin explorar la correlació entre certs moviments i les característiques de la parla en vídeos d'entrenament reals. Això ajudarà en la tasca d'aplicar models de predicció per crear o modificar animacions existents d'acord amb un àudio d'entrada. Ens hem enfocat en crear una eina modular i extensible per automatitzar els processos de capacitació i animació en lloc del desenvolupament real de models de predicció final, que requereixen conjunts de dades extensos, grans recursos informàtics i un profund coneixement de l'animació de personatges. Amb aquest projecte, hem aconseguit crear una eina modular i ampliable que automatitza el procés de capacitació i animació com a complement de *Blender*.

Castellano

En este documento describimos una plataforma para ayudar en la producción de uno de los aspectos que consumen más tiempo de la generación de imágenes por ordenador: la animación. Esta plataforma utiliza herramientas de vanguardia de múltiples campos como *Machine Learning* y *Computer Vision* para proporcionar un marco en el que los usuarios puedan explorar la correlación entre ciertos movimientos y las características del habla en videos de entrenamiento reales. Esto ayudará en la tarea de aplicar modelos de predicción para crear o modificar animaciones existentes de acuerdo con un audio de entrada. Nos hemos enfocado en crear una herramienta modular y extensible para automatizar los procesos de capacitación y animación en lugar del desarrollo real de modelos de predicción final, que requieren conjuntos de datos extensos, grandes recursos informáticos y un profundo conocimiento de la animación de personajes. Con este proyecto, hemos logrado crear una herramienta modular y ampliable que automatiza el proceso de capacitación y animación como complemento de *Blender*.

Contents

Part I
Definició de l'abast i contextualització

1 Introducció

1.1 Context

Recentment, el món de l'entreteniment ha estat revolucionat gràcies a les noves tecnologies (Smart TV's, millor connexió a Internet, hardware més potent), de forma que hi ha hagut un augment considerable del consum de sèries, pel·lícules [1] i videojocs [2] en els últims anys. Aquests tres medis, entre d'altres, tenen una peça en comú: les imatges generades per ordinador (o en anglès, Computer-Generated Imagery, CGI). Des del primer llargmetratge completament en CGI, *Toy Story* (1995), fins ara, ha passat de tenir un ús anecdòtic a ser una part essencial de la producció de contingut audiovisual. Observant aquesta tendència, podem deduir que la producció de CGI també seguirà augmentant.

Però, encara que cada cop hi ha més producció de CGI i millors eines per a realitzar aquest tipus de treballs, sembla que el cost no s'ha vist reduït amb el pas del temps [3]. Aquest fenomen pot ser degut a diverses raons: els projectes cada cop són més ambiciosos, les eines noves tenen cada cop més funcionalitats augmentant la complexitat, entre d'altres. Segons la llei d'Amdahl, la millora que podem obtenir al canviar un component d'un sistema està limitada per la fracció de temps que és utilitzat. En aquest context, per exemple, el cost per unitat de temps de crear un objecte 3d per a una pel·lícula es veu reduït a mesura que la pel·lícula augmenta de duració, ja que només cal crear-lo un cop. El cost d'animar un objecte 3d, però, és directament proporcional a la duració de la pel·lícula, com més llarga és, més treball necessita. Per tant, si volem reduir costos, seria sensat intentar reduir el cost d'aquelles tasques que són directament proporcionals, com ho seria l'animació.

La reducció dels costos de producció de CGI seria una de les aplicacions directes de l'eina que es presenta en aquest treball. **Speech2Anim** pretén ser capaç de proporcionar als usuaris una eina per explorar i desenvolupar models per modificar una animació preestablerta per a un avatar de forma que, utilitzant com a entrada el so d'una veu, aquest adopti una animació que s'adapti a la veu de forma versemblant. Aquesta eina podria reduir de forma significativa el cost d'animar un avatar que parla, proveint una base sobre la qual els animadors podrien treballar, generant animacions per avatars a un videojoc o bé generant animacions per a assistents virtuals.

1.1.1 Actors implicats

- **Director del projecte** És l'encarregat de guiar i aconsellar a l'autor d'aquest treball. Així com de supervisar el treball per garantir que compleix amb els requisits establerts dins dels terminis.
- **Autor del projecte** L'autor investiga, analitza, formula, resol i documenta el problema descrit en aquest document amb l'ajuda del director.
- **Usuaris finals** Els usuaris finals seran aquells que es vegin beneficiats per l'ús de l'eina que es proposa en aquest treball un cop acabat i funcionant. Aquesta eina podria ser útil als següents usuaris:
  - **Animadors professionals**: L'eina podria proporcionar una base per animacions més complexes, o animacions per altres personatges secundaris de fons.
  - **Programadors de videojocs**: Es podria fer servir aquesta eina per crear *placeholders* o crear contingut de forma massiva.
  - **Dissenyadors gràfics o audiovisuals**: L'eina podria ser útil per generar avatars per a narracions a presentacions o vídeos on no es disposi de temps o experts en animació.

1.2 Estat de l'art

En aquest apartat s'explora l'estat de l'art per als diferents camps relacionats amb el projecte. Alguns dels conceptes claus utilitzats per a fer aquest estudi són: *Affective Computing*, *Speech Recognition*, *Audio Analysis*, *Human Pose Detection*, *Neural Networks*, *Deep Learning*. I algunes de les conferències rellevants són: *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, *Affective Computing and Intelligent Interaction (ACII)*, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

1.2.1 Extracció de característiques audiovisuals

L'extracció de característiques serveix per conèixer quan succeeixen esdeveniments i de quina forma. Les eines actuals permeten extreure informació de diferents tipus, com ara: to, pauses, ritme, color, posició, etc.

Per a l'anàlisi de característiques acústiques es poden trobar eines com *openSMILE* [4] i *pyAudioAnalisis* [5] que permeten preprocessar el senyal, processar les dades per tal de diferenciar les característiques, extreure característiques del so de baix nivell (energia, espectre, volum, qualitat), detecció de punts importants (valors extrems, mitges, pics), detecció bàsica de característiques de vídeo (histogrames, detecció facial) i funcions de classificació, segmentació i regressió de les dades. Podem trobar estudis relacionats recents al *Computational Paralinguistics Challenge* (ComParE) [6], un repte que consisteix a determinar trets i estats dels parlants basant-se en les propietats acústiques de la parla (Per exemple [7]).

Mentre que habitualment l'àudio i la imatge s'analitzen per separat, estudis recents utilitzen l'anàlisi conjunt per trobar esdeveniments de forma més precisa [8].

1.2.2 Reconeixement de la parla

El reconeixement de la parla és el camp que es centra a desenvolupar tecnologies capaces de entendre la parla i traduir-la a format de text. Això permet a un sistema mostrar, guardar o processar el text per a entendre el context.

Trobem eines com la *API Speech* de *Google Cloud Platform* [9], *Speechmatics* [10] i el *English Conversational Telephone Speech Recognition System* d'*IBM* [11] que, d'entre altres coses, poden fer el reconeixement en diversos llenguatges, en directe i amb tractament del soroll.

Un dels reptes més significatius d'aquest camp és solucionar el *Cocktail Party Problem*, que consisteix en aconseguir el reconeixement de la parla per a una font en concret, filtrant tota la resta, ja siguin soroll o altres parlants. Existeixen diversos intents de solucionar aquest problema utilitzant diferents tecnologies [12] [13].

1.2.3 Extracció de postures humanes

Fa pocs anys era habitual l'extracció de postures utilitzant *Depth Frames* (imatges de fondària) que va posar de moda *Microsoft* amb la comercialització de la *Kinect*, una càmera amb sensors de fondària creada expressament per a aquesta tasca [14].

Mentre que la *Kinect* estimava les postures del cos i el cap, actualment existeixen eines com *OpenPose* [15] [16] [17] que poden, a més, estimar la posició de les mans, els dits i les expressions facials, tot a partir d'imatges de color.

Depenent de l'ús, existeixen mètodes amb diferents característiques. Els que van produir millors resultats al *Pascal VOC Challenge* [18] van ser els que utilitzaven *Deformable Parts Models* [19]. A l'estudi [20], es millora la velocitat d'un mètode de detecció amb pèrdua de precisió negligible utilitzant *Aggregate Channel Features*, permetent el càlcul en temps real. A [21], desenvolupen un procés anomenat *Iterative Error Feedback* (IEF) que fa una estimació de la postura realitzant múltiples iteracions corregint errors a cada passada, aconseguint uns resultats més precisos que altres mètodes.

1.2.4 De veu a animació

Existeixen diversos estudis que contemplen les possibilitats de crear animació basada en veu com a entrada. Una aplicació bastant directa és la de generar el moviment de la boca mitjançant una anàlisi dels fonemes [22] [23]. A altres estudis també es genera animació facial [24] [25], tot i que estan enfocats la comunicació en temps real i, en alguns casos, necessiten altres entrades addicionals com una càmera web. En el cas de [26], s'aconsegueixen expressions facials, sincronització dels llavis i moviments del cap d'acord amb el contingut emocional de l'àudio d'entrada.

A [27], es té en compte el significat de la parla per a produir animacions facials i, a [28], a més, es capturen les relacions que existeixen entre els diferents músculs i moviments de la cara.

Pel que fa a postures, es poden trobar estudis que prediuen la posició del cap [29] segons la veu. A [30], es generen animacions que inclouen expressions facials, gestos i moviments de les mans i els braços en relació a la parla basant-se en un sistema de generació de regles. En aquest cas, però, no es generen a partir de la parla, sinó que la parla també és generada pel mateix agent. Per últim, a [31], es proposa un sistema que basant-se en la parla, pot generar gestos i animacions (mans i cap) sincronitzades amb l'àudio.

1.2.5 Conclusions

Dels apartats anteriors podem concloure que tots aquests camps estan en ple desenvolupament. Alguns dels estudis citats són tan recents com d'aquest mateix any. A més a més, són tecnologies que encara no són del tot estables o acurades a hores d'ara.

Donat que els algoritmes darrere de les eines d'anàlisi d'àudio i vídeo són molt complicats, no es proposaran millores en aquest aspecte. Aquest treball busca explorar la capacitat útil d'aquestes noves tecnologies creant una eina basada en elles.

2 Formulació

2.1 Problema

Podem dividir el problema en dos parts:

- **Comunicació:** A l'hora d'entendre un discurs o una conversació, el llenguatge no verbal juga un paper molt important [32]. Les persones utilitzen la cara, el cap i les mans per donar èmfasis a parts del missatge. Quan un missatge és escoltat sense llenguatge corporal, es redueix la seva comprensió [33].
- **Creació de contingut:** Animar des de zero un avatar que parla pot ser una tasca molt costosa. En el cas dels videojocs on es va afegint contingut de forma continuada, pot ser del tot inviable.

2.2 Abast

En aquest treball es proposa un sistema per modificar animacions existents introduint canvis de posició i/o velocitat per a algunes parts del cos amb la intenció de solucionar els problemes descrits. Per tal de:

- Generar una animació que serveixi com a base per a aconseguir animacions més detallades, estalviant temps i cost.

- Aconseguir animacions bàsiques per a utilitats més secundàries o amb contingut massiu, com per exemple videojocs.
- Aconseguir animacions que serveixin per a presentacions o altres aplicacions on no es disposi d'un artista.
- Aconseguir que les animacions generades reforcin el missatge, ajudant a fer que s'entengui millor.

Aquest treball no pretén aconseguir el següent:

- Generar l'animació en temps real.
- Generar l'animació sense cap tipus de preparació o entrenament.
- Generar l'animació de zero, sense ajuda d'altres animacions generades per artistes o capturades.

2.2.1 Objectius

- Aconseguir extreure informació dels vídeos d'entrenament en forma de postures humanes, animacions, velocitats, etc.
- Aconseguir crear un model que predigui la informació necessària per modificar posicions i velocitats per a algunes parts de l'esquelet a una animació existent.
- Aconseguir que les modificacions generades donin una sensació versemblant de que l'avatar està dient el mateix que a l'àudio d'entrada.

2.2.2 Possibles obstacles

Hi ha tota una sèrie de punts que podrien obstaculitzar el desenvolupament d'aquest projecte. Per començar, els models a crear depenen completament de la informació que es pugui extreure mitjançant eines ja existents. Si aquesta informació resultés ser insuficient, imprecisa o incorrecte, es dificultaria molt aconseguir un resultat satisfactori.

Un altre punt és que l'existència de models que puguin explicar les postures és una suposició optimista (basada en evidència científica, com s'ha explicat anteriorment) per part del director i l'autor del projecte. Es desconeix si es trobaran aquests models o no. Per això, l'èmfasi serà crear l'eina que permeti l'experimentació amb eines de visualització per veure les dades audiovisuals i els senyals de forma gràfica, més que el desenvolupament de models finals pròpiament dits, que requereixen conjunts de dades i recursos de computació fóra de l'abast d'aquest projecte.

A més, encara que es solucionin aquests problemes, el resultat del projecte depèn en gran mesura de la percepció subjectiva dels usuaris. Podria passar que s'obtinguessin uns resultats precisos estadísticament, però que no fossin agradables de veure. Aquest fenomen es coneix com a *La Vall Inquietant* (*The Uncanny Valley*) [34].

Finalment, podrien donar-se altres problemes més comuns. En fer ús de tecnologies noves, poden sorgir errors o incompatibilitats inesperades. Es pot esperar una comunitat menor a Internet i, per tant, menys suport.

3 Metodologia

3.1 Mètode de treball

A causa de la naturalesa del projecte, s'ha decidit adoptar una metodologia de treball àgil. En concret, la metodologia *Kanban*. Aquest mètode permet reaccionar ràpidament a canvis i imprevistos, progressa seguint la capacitat del desenvolupador en comptes de cicles preestablerts i facilita estar centrat en les tasques importants sense sacrificar la percepció global del projecte. A més, existeixen eines en línia gratuïtes que en permeten la gestió.

Kanban consisteix en un tauler visual que disposa de diverses seccions, on s'afegeixen targetes que representen tasques. Aquestes seccions indiquen en quin estat es troben les tasques que contenen. Per tant, donant un cop d'ull al taulell, es pot saber en quin estat es troba el projecte en general i cada tasca concretament. Essencialment existeixen dos rols, el propietari i el desenvolupador. El propietari s'encarrega d'afegir i ordenar les targetes, mentre que el desenvolupador agafa les targetes que li pertoquen en la mesura que pot. Un dels punts fonamentals és mantenir les tasques en un ordre concret, que permeti el progrés correcte del projecte.

3.2 Eines de seguiment

Per mantenir un seguiment del projecte, s'utilitzarà l'eina **Trello**, que implementa el mètode *Kanban* abans descrit, amb altres característiques addicionals.

També s'utilitzarà un repositori **git** allotjat a **Github**. El repositori aporta diversos avantatges: historial dels canvis realitzats, capacitat de desfer canvis, opció de fer proves utilitzant *branches* secundàries i còpia de seguretat allotjada a Internet.

3.3 Mètode de validació

Per tal de validar el progrés, definirem algunes de les tasques com a *milestones*, de forma que s'han d'anar assolint a una certa data per comprovar que s'avança a un ritme adequat. A més, es realitzaran reunions periòdiques amb el director per rebre *feedback* i orientar el projecte.

Part II

Teoria

4 Extracció de característiques

4.1 Imatges i postures humanes

Per a entrenar els models, cal extreure informació de les postures dels vídeos utilitzats als entrenaments. En aquest projecte s'ha utilitzat l'eina *OpenPose* [15] [16] [17], que utilitza *Convolutional Pose Machines* i *Part Affinity Fields* per identificar postures de múltiples persones a una sola imatge i a temps real.

4.1.1 *Convolutional Pose Machines*

Les CPM (*Convolutional Pose Machines*) consisteixen en una sèrie de predictors (*convolutional networks*) entrenats per analitzar diferents parts d'una imatge de forma seqüencial. Cada predictor s'associa amb una part del cos, de forma que a cada iteració genera un *belief map* (mapa de confiança) que indica a on creu que es troba aquella part del cos concreta. Aquests mapes serveixen per expressar la incertesa espacial de les diferents parts del cos.

Figura. 1: Mapes de confiança de la posició del canell dret.
(a) Confusió deguda a informació exclusivament local. (b) Context de diverses parts ajuden a reduir ambigüitat. (c) Iteracions addicionals ajuden a convergir a una solució.

A cada iteració, cada predictor es basa en la resta de mapes de confiança generats pels altres predictors per refinar el mapa següent. D'aquesta forma, a cada iteració es milloren les estimacions de cada part del cos (vegeu Figura 1).

4.1.2 *Part Affinity Fields*

Els mapes de confiança generats per les CPM tenen la limitació d'indicar només la posició de les parts del cos. Això provoca que, en imatges on existeixen diverses persones, sigui difícil dir quines de les parts formen part del mateix individu. Els PAFs (*Part Affinity Fields*) solucionen aquest problema afegint a les parts una estimació de la seva orientació.

Figura. 2: Esquerra: PAFs definits pel colze i el canell drets per a múltiples persones. El color representa l'orientació. Dreta: Vectors que representen l'orientació de cada punt.

Per generar els PAFs, s'assigna a cada punt de la imatge un valor, que o bé, és el vector unitari format pels dos punts que representen els extrems d'una part del cos (per exemple el colze i el canell) o 0 si no forma part de cap part del cos (vegeu Figura 2). Aquest vector el podem trobar a partir dels punts obtinguts als mapes de confiança.

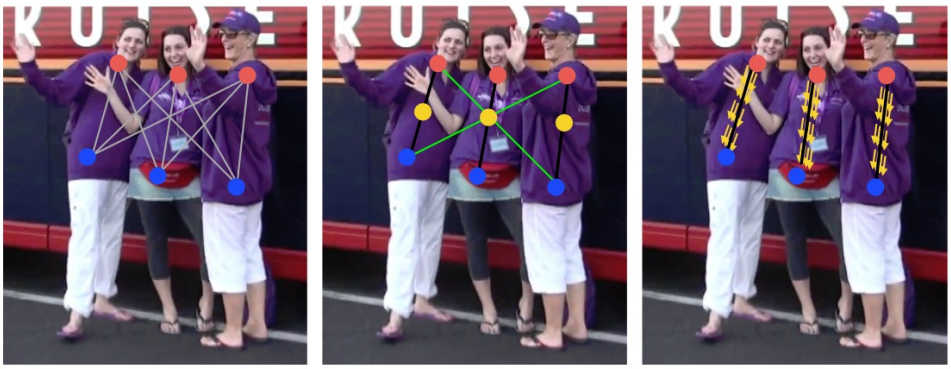


Figura. 3: (a) Graf d'associacions (b) En groc els punts mitjans generats. En verd les associacions incorrectes (c) Mitjançant PAFs es pot escollir l'associació correcte, sense ambigüitat en aquest cas.

A l'hora d'associar parts del cos, es construeix un graf bipartit on els punts són parts del cos i les arestes possibles associacions (vegeu Figura 3a). Una possible forma de decidir quines associacions són correctes, seria generar un punt mitjà per cada possible associació i agafar les arestes que hi passen per allà, però això pot provocar altres errors (vegeu Figura 3b). Quan es troben diferents persones juntes a una imatge, els PAFs ajuden a oferir el context necessari per escollir les associacions correctes (vegeu Figura 3c).

4.2 Àudio

Per obtenir característiques de l'àudio s'ha utilitzat l'eina *openSMILE* [4], un *framework* extensible i modular que permet extreure descriptors de l'àudio i vídeo, preprocessar el senyal i postprocessar el resultat per tal construir resums o exportar les dades en un format concret. L'arquitectura es basa en 4 elements principals:

- ***Data Memory***: Emmagatzematge de qualsevol tipus de dada.
- ***Data Sources***: Components que llegeixen dades de fonts externes.
- ***Data Processors***: Components que copien i modifiquen les dades.
- ***Data Sinks***: Components que escriuen les dades a destins externs.

Utilitzant aquest tipus de components, es poden construir uns arxius de configuració que especifiquin quines dades s'han d'extreure de l'àudio, com s'han de processar i com s'han de guardar. A més, aquests arxius permeten utilitzar definicions de constants i variables que es poden modificar mitjançant paràmetres d'entrada. A continuació una breu llista de les característiques rellevants per a aquest projecte:

- **Processament d'àudio**: *Windowing Functions, Fast-Fourier Transform (FFT), filtering.*
- **Extracció de *features* relacionades amb la parla (*Low-level descriptors*, LLD)**: *Signal Energy, Loudness, Pitch, Perceptual Linear Prediction* (PLP), *Linear Predictive Coding* (LPC), *Mel-Frequency Cepstral Coefficients* (MFCC).
- **Funcions estadístiques (*Functionals*)**: mitjanes, extrems, pics, *Zero-crossings.*

En aquest treball, s'han fet servir les funcions estadístiques d'un fitxer de configuració existent *ComParE_2016* modificat per acceptar certs paràmetres.

4.2.1 *Linear Predictive Coding*

LPC es basa en aproximar una mostra $s[n]$ de veu en base a p combinacions lineals de mostres anteriors $s[n-1], s[n-2], ..., s[n-M]$ més un senyal d'error $e[n]$:

$$s[n] = \sum_{i=1}^{M} a_i s[n-i]$$

on $a_i, i = 1, 2...M$ són els predictors, anomenats *LPC coefficients*. La clau del LPC és el fet que podem trobar el senyal d'error $e[n]$ aplicant un filtre invers sobre $s[n]$, i, per tant, podem aplicar el criteri dels quadrats mínims per minimitzar-la.

4.2.2 *Mel-Frequency Cepstral Coefficients*

Els anomenats *Mel-Frequency Cepstral Coefficients* són *features* que s'extreuen d'àudios amb contingut de parla que serveixen per augmentar la informació a l'hora d'entrenar models que la reconeguin. Els MFCC, són molt populars per a tasques de processament d'àudio i detecció de *features*.

Per computar els MFCC s'han de realitzar les següents tasques [35]:

- Separar els *frames* en grups (finestres).
- Obtenir el FFT del senyal per a una finestra per obtenir la descomposició de les freqüències del senyal.
- Mitjançant finestres triangulars sobreposades, obtenir les freqüències del pas anterior i obtenir els valors de l'escala de Mel corresponents.
- Calcular els logaritmes a cadascun dels valors de Mel obtinguts.
- Calcular el *Discrete Cosine Transform* sobre el conjunt de logaritmes calculats al pas anterior.
- Els MFCCs són les amplituds de l'espectre resultant.

4.2.3 *Perceptual Linear Prediction Cepstral Coefficients*

La PLP és molt similar als MFCC, però amb la diferència principal que combina components psicofísics de l'oïda:

- **Selectivitat espectral de la banda crítica**: L'amplada de banda d'un soroll a partir de la qual ja no augmenta l'efecte d'emmascarament.
- **Corbes isofòniques**: Corbes que calculen la relació entre la freqüència i el volum percebut.
- **Llei de l'energia de Stevens**: Una proposició de la relació que hi ha entre la magnitud d'un estímul físic i la força o intensitat amb la qual es percep.

5 *Machine Learning*

Per tal d'entrenar els models utilitzats per a generar les animacions, es necessiten algorismes d'aprenentatge. Per a aquesta fi, s'ha fet ús de la GRT (*Gesture Recognition Toolkit*), una llibreria de *Machine Learning* dissenyada específicament per a la detecció de gesticulacions. Aquesta, conté diversos algorismes de classificació que s'han fet servir per generar els models. De tots els models generats, s'ha fet servir *Matthew's correlation coefficient* per seleccionar el millor, el qual dóna un valor entre -1 i +1 que indica la correlació entre les classificacions observades i les classificacions predites. -1 indica desacord total, mentre que +1 indica predicció perfecta.

Aquest projecte permet als usuaris afegir vídeos i característiques per tal de construir nous models. Per aquesta raó, és important que aquest sigui un procés automàtic i pugui servir per entrades potencialment diferents entre si.

5.1 *Datasets*

Per produir els models per defecte s'han emprat diverses fonts d'informació. Inicialment, es van utilitzar vídeos del *dataset First Impressions* [36]. Seguidament, es van fer servir vídeos de llarga duració de la mateixa persona i diferents perfils (*talks*, normal, etc). A l'hora de seleccionar-los s'ha tingut en compte que gesticulessin adequadament i que es veiessin bé de cintura cap a dalt. Així com la absència de més gent a la imatge i un punt de vista estable al llarg del vídeo (o fàcilment editable). Per al model final s'han fet servir extractes dels següents vídeos:

- *Why Sitting Down Destroys You - Roger Frampton- TEDxLeamingtonSpa* [37].
- *Intervención de Albiol PP ante el Parlamento de Cataluña 10 Octubre 2017* [38].

5.2 Algorismes

Després de realitzar diverses proves amb diferents algorismes i paràmetres, s'ha descobert que, en general, aquests algorismes trobaven els millors models.

5.2.1 KNN

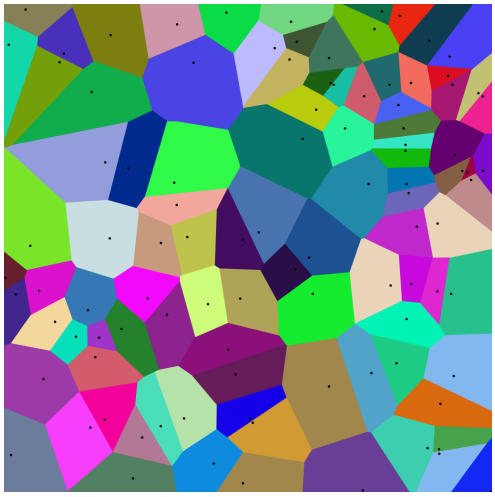


Figura. 4: Distribució de l'espai de classes utilitzant un valor de k=1. Cada color representa una classe, per tant, cada punt té una classe diferent.

K nearest neighbours es un algorisme senzill de classificació. Consisteix en, donat un *sample* s, trobar els K veïns més propers a s i classificar-lo amb la mateixa etiqueta.

D'aquesta forma, amb $K = 1$, es busca el *sample* més semblant a la mostra i retorna la seva etiqueta. Per $K > 1$, es busquen els K elements més propers i es retorna l'etiqueta més repetida. És, per tant, un *lazy learner*. En comptes de processar les dades d'entrenament, simplement les guarda per comparar-les amb els *samples* d'entrada. Això fa que sigui lent a l'hora de fer prediccions, sobretot quan hi ha un *training set* gran.

Sabent això hem de tenir en compte el següent:

- Si tenim un problema de classificació binaria, haurem d'escollir una K senar per evitar que hi hagi empats d'etiquetes.
- De la mateixa forma, haurem d'escollir preferiblement valors de K de forma que no siguin múltiples amb el nombre de classes.
- És lent per a *training sets* grans.
- És susceptible al soroll.

5.2.2 MinDist

Aquest algorisme, a diferència del *KNN*, sí que aprèn a partir de les dades d'entrenament. Accepta un paràmetre M que especifica un número de *clusters* en els quals es classifiquen les dades d'entrenament, basant-se en la distància euclidiana entre les mostres. A l'introduir un *sample* d'entrada, només cal calcular la distància als clústers i tornar l'etiqueta del més proper.

Sabent això hem de tenir en compte el següent:

- El resultat pot variar molt depenent del paràmetre M que escollim, per tant és important provar amb diferents valors.
- Hi han d'haver almenys tants *clusters* com classes.
- És més ràpid que *KNN*.

Part III

Disseny

Des de bon començament, es va decidir implementar el projecte en forma de *add-on* per *Blender*, un programa *Open Source* d'edició 3D que disposa d'un intèrpret de *Python* per automatitzar tasques. *Blender* disposa d'una varietat molt amplia de eines que podem aprofitar per desenvolupar aquest projecte i, a més, d'una bona documentació i comunitat. Es va decidir programar el projecte en *Python* no només per la compatibilitat amb *Blender*, sinó per la seva flexibilitat. Com veurem més endavant, això permetrà un disseny molt extensible i fàcil de programar.

Com que s'ha utilitzat *GRT* per generar els models, també s'ha desenvolupat una part del projecte amb *C++*. Això comporta un inconvenient i un avantatge. Per una banda, es dificulta la portabilitat, és a dir, s'ha de generar un executable per cada plataforma. Però, per l'altre, permet substituir la part de generació de models intercanviant l'executable per un altre que pugui interpretar l'entrada generada.

Ja que es preveia que la part d'entrenament i animació seria molt exploratòria, basada en experimentar amb diferents parts del cos (cap, braços), aspectes de l'animació (posició, orientació, velocitat, patrons de moviment), es va decidir posar més pes en desenvolupar una plataforma que facilités aquesta feina en comptes de proporcionar una solució rígida. D'aquesta forma, donem control a l'usuari perquè esculli amb quina informació entrenar el model i com interpretar-la, permetent utilitzar diferents configuracions per a diferents projectes i aconseguint una major expressió.

Per exemple, un usuari pot voler estudiar correlacions entre la velocitat angular del cap, mesurada prenent com a referència el nas i el coll, en una finestra de 500 ms, amb característiques de l'àudio. Més endavant li pot interessar detectar quan l'usuari desvia notablement la mirada respecte la càmera, o es porta les mans al cap. En quant a la part de síntesi d'animacions plausibles, l'usuari pot desitjar introduir una animació procedural determinada, o carregar una animació predefinida només quan es prediuen determinades accions o esdeveniments.

6 Eines externes i estructura

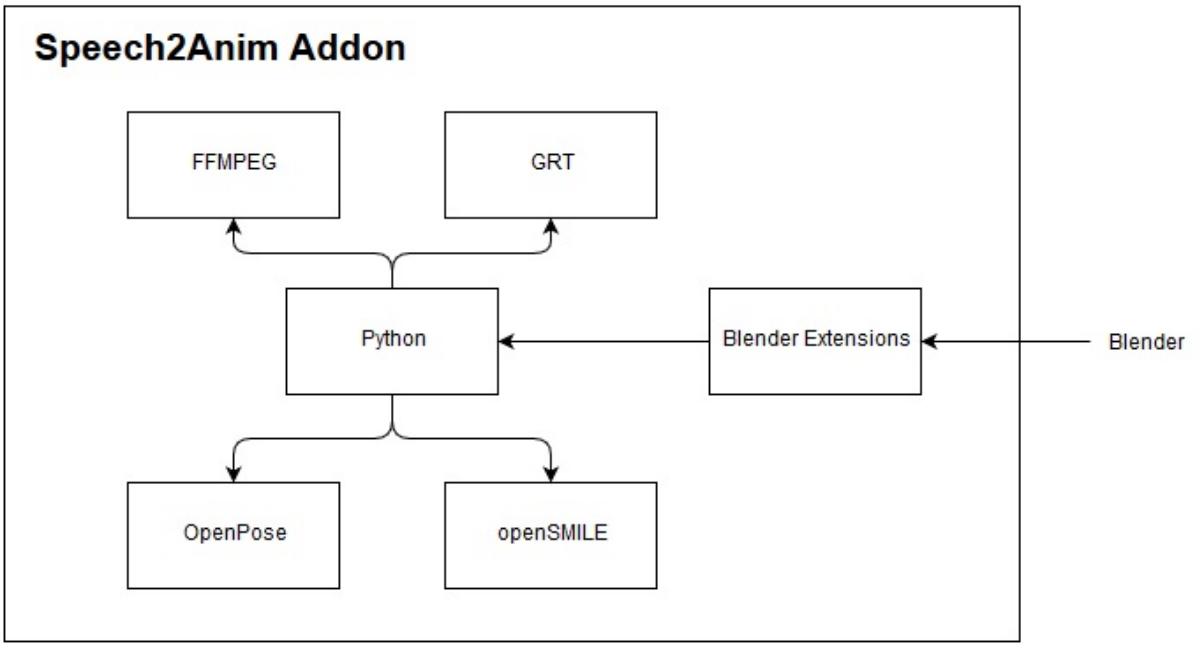


Figura. 5: Estructura de l'*add-on*. Les fletxes indiquen '*crida a*'.

L'estructura del projecte és simple: unes extensions de *Blender* que exposen la funcionalitat dels *scripts* de *Python* que interactuen amb llibreries de tercers. A continuació es troba una breu descripció de cada part:

- ***Blender Extensions***: Una sèrie de classes que hereten de classes bàsiques de *Blender* que serveixen per estendre la seva funcionalitat, com, per exemple, afegint una nova finestra o una nova estructura de dades.
- ***Python***: *Scripts* que s'encarreguen de transformar i moure les dades i executar les diferents eines necessàries a cada pas.
- **GRT**: Executable preparat per generar models i prediccions fent servir la llibreria GRT.
- **FFMPEG**: *Software* que permet editar i convertir fitxers d'àudio i vídeo. En aquest cas s'utilitza per extreure l'àudio dels vídeos en un format concret.
- ***OpenPose***: *Software* que permet extreure informació en 2D sobre les postures de les persones que surten a un vídeo. S'utilitza a l'hora de generar les dades d'entrenament.
- ***openSMILE***: *Software* que permet extreure una gran quantitat de *features* de l'àudio. S'utilitza tant com per generar les dades d'entrenament com per generar l'*input* de les prediccions.

7 Mòduls

En aquest apartat es pot apreciar amb més detall l'estructura de la part de *Python*.

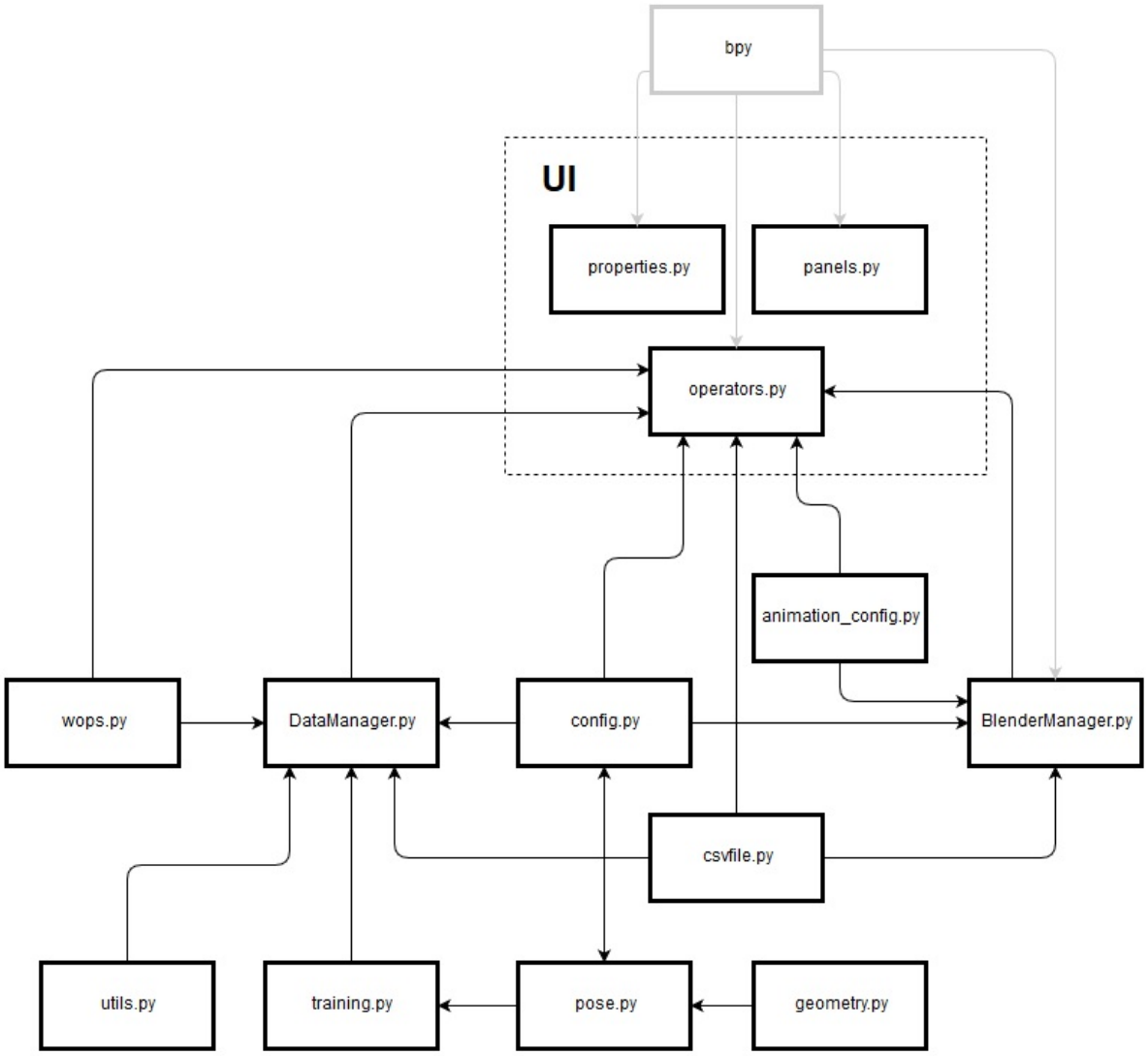


Figura. 6: Diagrama de mòduls. La direcció de les fletxes significa '*Importat per*'. S'han omès els mòduls estàndard. S'ha afegit el mòdul de *Blender*, *bpy*, en color gris clar. El quadre de línia discontinua (UI) agrupa els mòduls de la interfície gràfica.

Al diagrama anterior (6) podem veure una descripció de la relació entre els diferents mòduls implicats. A continuació es troba una llista amb una breu descripció de cada mòdul:

- **DataManager.py**: Defineix la lògica principal del programa. Utilitza la resta de mòduls i eines per a dur a terme les tasques d'entrenar i predir. Per motius històrics, ha conservat (tot i que no s'utilitza) un *main* i accepta paràmetres d'entrada.

- **BlenderManager.py**: Defineix alguns mètodes que s'utilitzen per importar les dades dins de *Blender*
- **config.py**: Defineix les constants (*paths*, valors numèrics, noms) que s'utilitzen al programa i les funcions utilitzades per processar i generar les *features* d'entrenament. Aquest mòdul pot ser proporcionat per l'usuari i està dissenyat per tal que sigui fàcil la seva edició.
- **animation_config.py**: Defineix les animacions que interpreten les prediccions del model. Aquest mòdul pot ser proporcionat per l'usuari i està dissenyat per tal que sigui fàcil la seva edició.
- **training.py**: Defineix les funcions principals que es fan servir per generar les dades d'entrenament.
- **csvfile.py**: Un petit mòdul amb utilitats per gestionar fitxers *csv*.
- **wops.py**: Conté mètodes per interactuar amb el sistema de fitxers. S'anomena *wops* (*Windows OPerationS*)perquè, al principi, estava implementat amb comandes de *Windows*. Després es va descobrir el mòdul *shutils* i es va utilitzar per a reimplementar el mòdul, tot i que es van mantenir els noms.
- **pose.py**: Defineix estructures de dades per tractar les postures estretes dels vídeos.
- **geometry.py**: Defineix classes per tractar punts i vectors.
- **utils.py**: Defineix altres funcions útils.
- **bpy**: Accés a l'estat i a les funcionalitats de *Blender*.
- **UI**:
  - **properties.py**: Defineix les estructures de dades necessàries per mantenir l'estat de l'*add-on*.
  - **operators.py**: Defineix les accions que afegeix l'*add-on*. Aquests operadors són l'enllaç entre *Blender* i el codi principal del projecte.
  - **panels.py**: Defineix la interfície gràfica exposant propietats i operadors.

7.1 Diagrames de flux de dades

Per entendre millor el funcionament de l'*add-on*, aquí es mostren dos diagrames de flux de dades per a les funcionalitats principals: l'entrenament d'un model i la generació d'una animació.

7.1.1 Entrenament

En aquest diagrama de flux es mostra el procés d'entrenament.

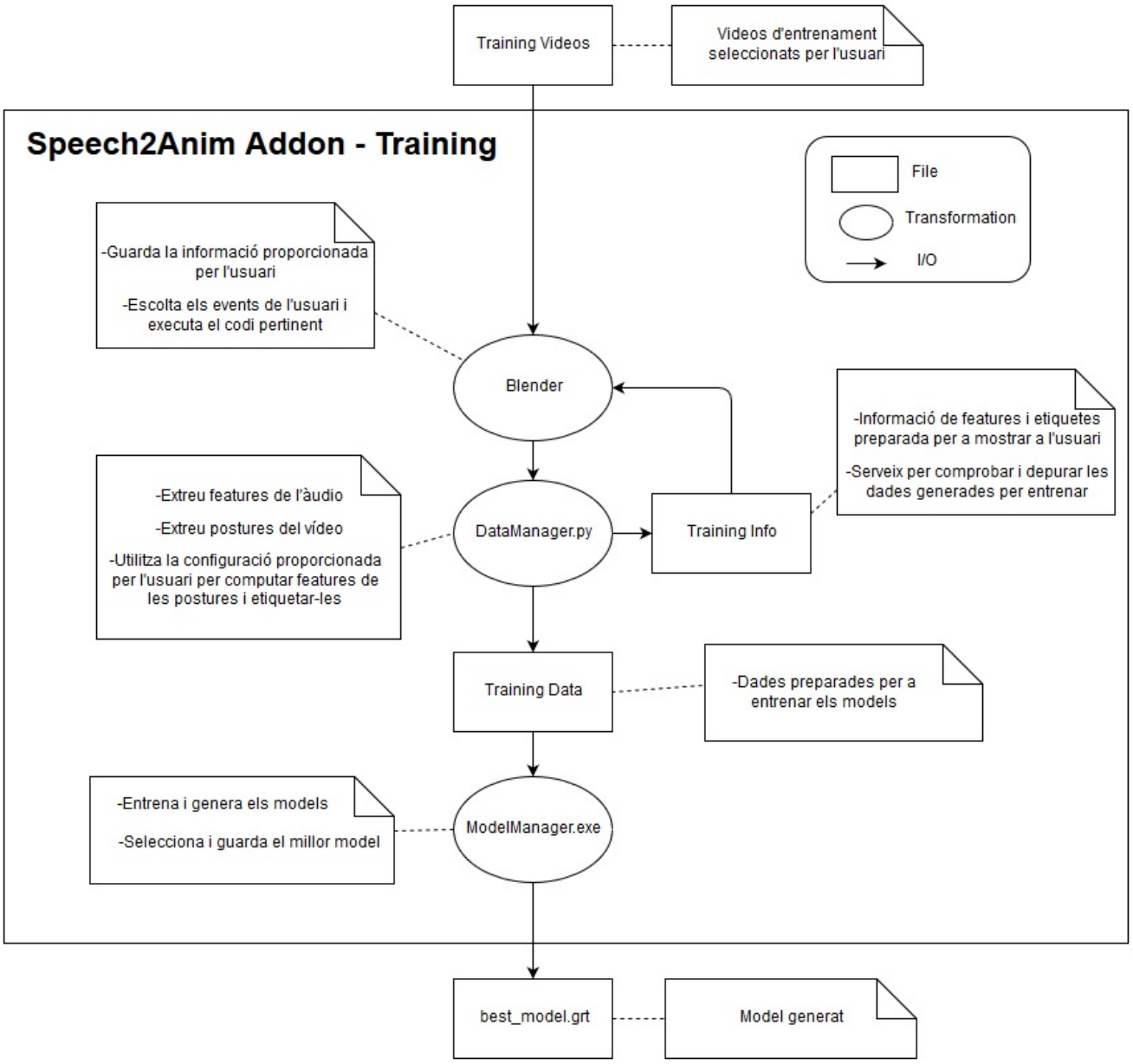


Figura. 7: Diagrama de flux de dades de l'entrenament d'un model.

7.1.2 Generació d'animació

En aquest diagrama de flux es mostra la generació d'animació.

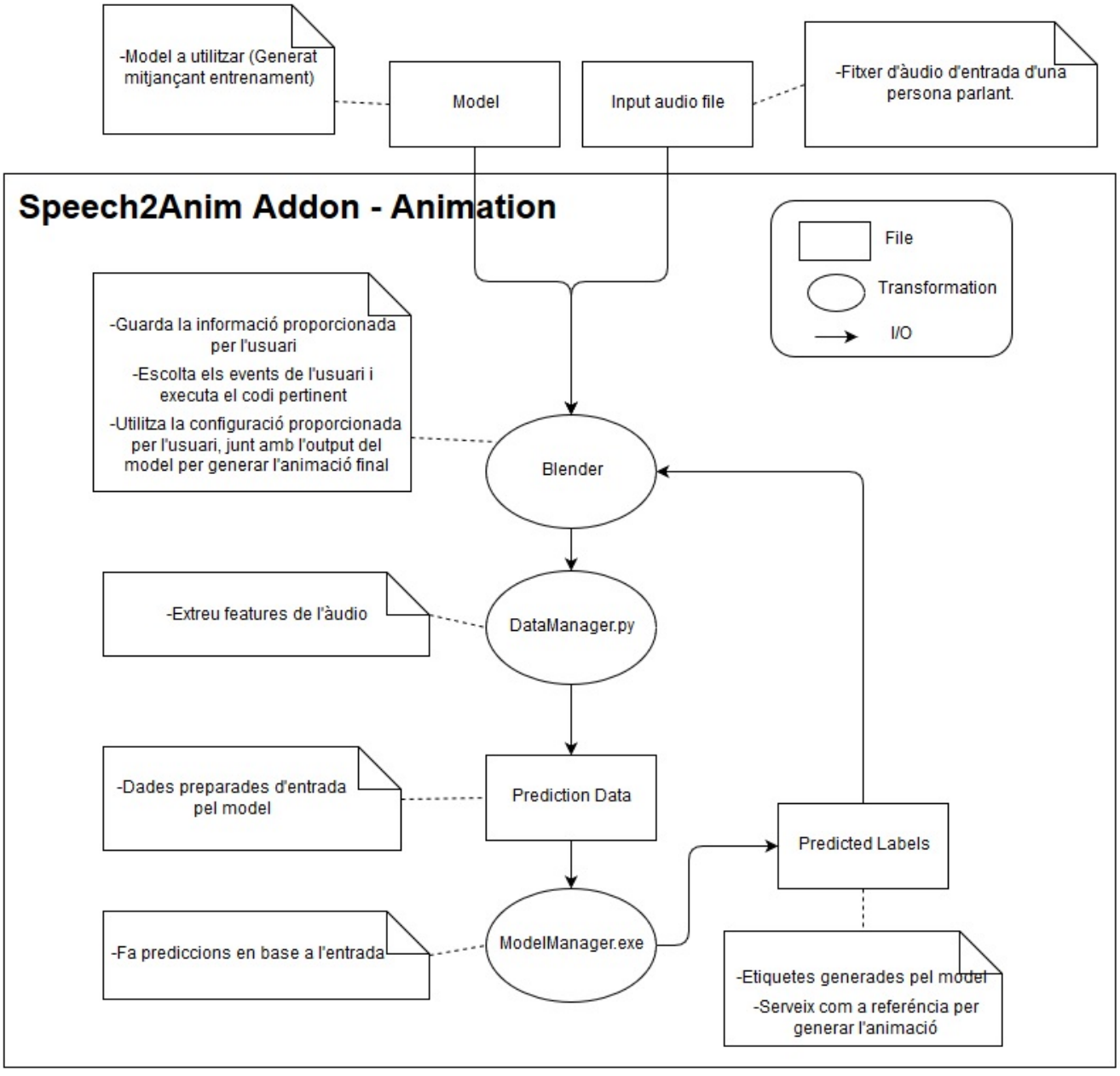


Figura. 8: Diagrama de flux de dades de generació de l'animació utilitzant un àudio i un model d'entrada.

7.2 Interfície

El *add-on*, en si mateix, registra dos menús, dins de *Blender*, ubicats dins el panell de propietats d'objecte

7.2.1 Menú d'entrenament

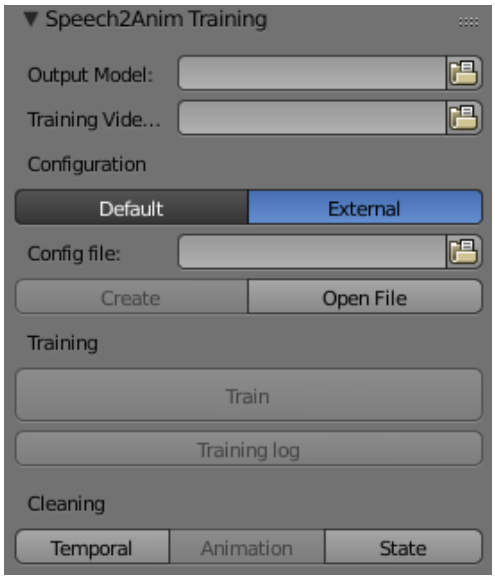


Figura. 9: Panell d'entrenament

A continuació es detalla l'ús dels controls que apareixen a la imatge (9).

- ***Output model***: Camí del fitxer on es guardarà el model generat.
- ***Taining videos folder***: Camí a la carpeta on es troben els vídeos d'entrenament.
- ***Configuration***: Podem escollir entre la configuració per defecte (*default*) o utilitzar una creada per l'usuari (*external*).
- ***Training***: El botó *train* genera el model utilitzant els vídeos i configuració especificats. I el botó *training log* permet veure informació sobre el procés d'entrenament. Un cop s'utilitza apareix el menú de la figura (10)
- ***Cleaning***: En aquest apartat trobem botons per netejar els arxius temporals o la informació carregada.

Figura. 10: Canvi a la interfície després de l'entrenament

7.2.2 Menú d'animació

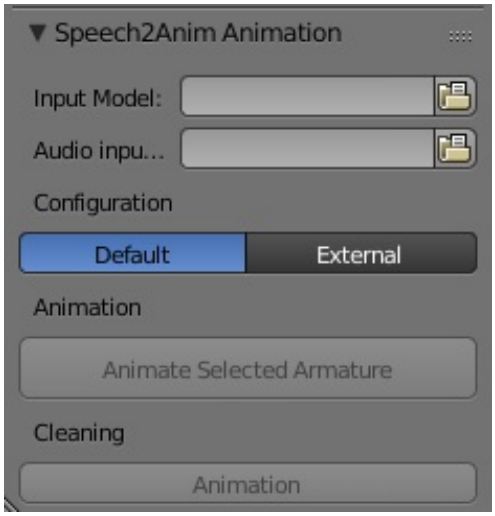


Figura. 11: Menú d'animació

A continuació es detalla l'ús dels controls que apareixen a la imatge (11).

- ***Input model***: Camí al model utilitzat per generar l'animació.
- ***Audio input file***: camí al fitxer de àudio que conté el discurs que es vol animar.
- ***Configuration***: Podem escollir entre la configuració per defecte (*default*) o utilitzar una creada per l'usuari (*external*).
- ***Animation***: El botó *animate selected armature* genera l'animació a l*armature* seleccionat.
- ***Cleaning***: El botó d'aquest apartat permet netejar l'animació del model.

7.2.3 Fitxer addicional

Junt amb el *add-on* es distribueix un fitxer *Blender* (.blend) que conté un model amb esquelet, animacions i dos *layouts* que faciliten el procés d'entrenament i animació.

7.2.4 *Layout* d'entrenament

El *Layout* d'entrenament es composa de les següents finestres:

- ***Video sequence editor***: En aquest espai es pot veure la detecció de postures realitzada per *OpenPose*.
- ***Text editor***: Aquest apartat resulta útil per editar les configuracions.

- ***Graph editor***: Aquí veurem els valors generats per cada *window value* i *functional*. El qual resulta d'utilitat per depurar el fitxer de configuració o explorar les dades d'entrenament.
- ***Timeline***: Permet desplaçar-se en el temps.

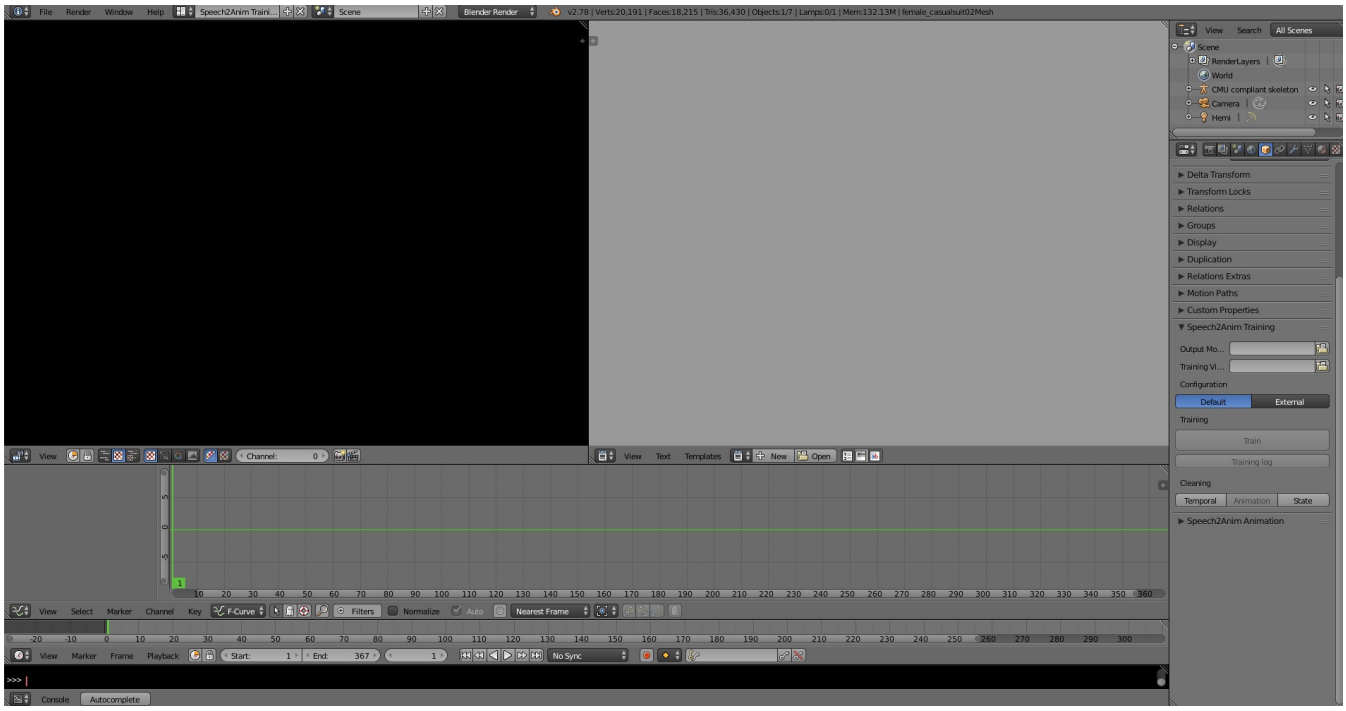


Figura. 12: *Layout* d'entrenament

7.2.5 *Layout* d'animació

El *Layout* d'animació es composa de les següents finestres:

- ***3D view***: En aquest finestra podem veure l'animació resultant.
- ***Text editor***: Aquest apartat resulta útil per editar les configuracions.
- **NLA *editor***: En el qual es veuen les composicions de les animacions realitzades.
- ***Timeline***: Permet desplaçar-se en el temps.

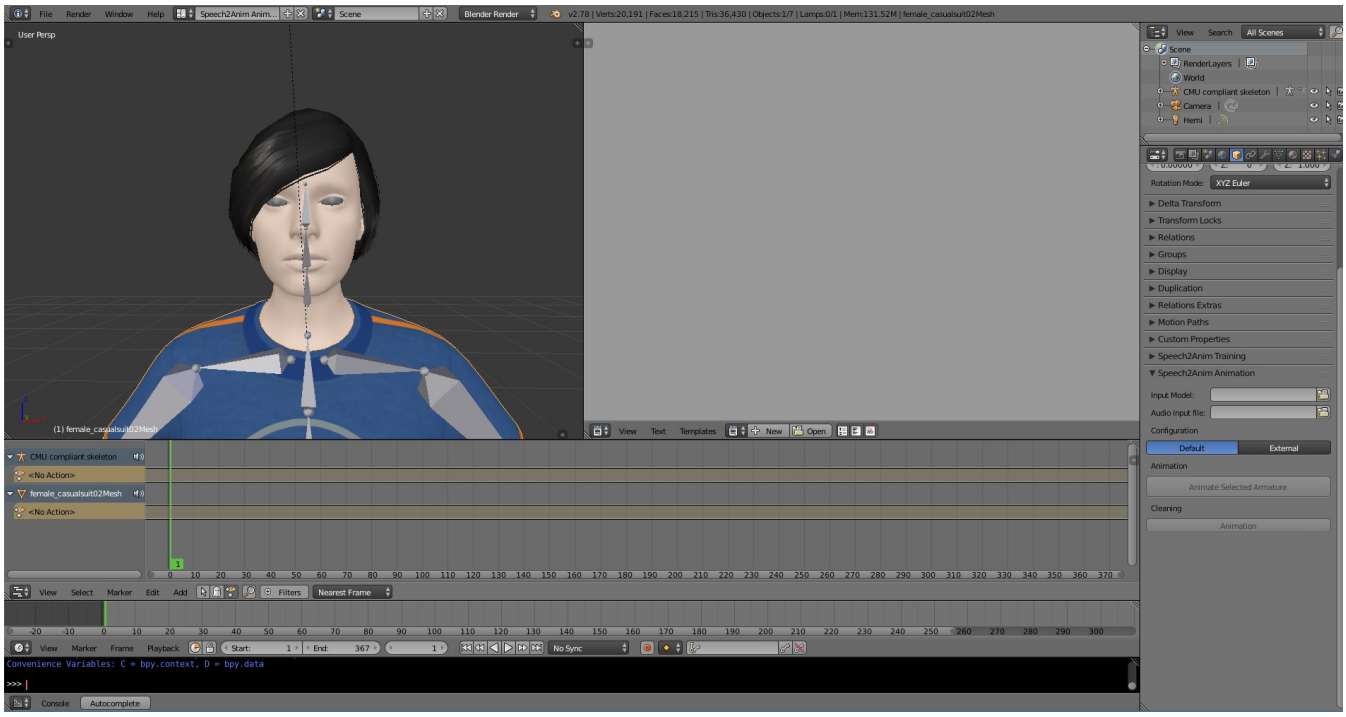


Figura. 13: *Layout* d'animació

7.3 Configuracions per defecte

7.3.1 Configuració d'entrenament

- Es poden definir els següents valors de configuració: La configuració que utilitza el *openSMILE* per extreure els *features* del àudio.
- La mida i el *step* de les finestres per calcular les *features* que s'extreuen del *OpenPose* i el *openSMILE*.
- ***Window values***: Es defineix un diccionari que conté les funcions que calculen o extreuen les *features* de la postura.
- ***Functional***: Es calcula un altre diccionari que conté funcions per resumir les definides en el diccionari anterior. Es defineix una llista de grups d'etiquetes (*labels*). Sent els labels de cada grup mútuament excloents entre si. Es generarà un model de cada grup.

7.3.2 Configuració d'animació

Es defineix la interpretació de la informació que prediu el model. En altres paraules, es transforma la classificació en animació. Per conveniència, el procés es realitza mitjançant una llista d'animacions que s'apliquen una darrera de l'altre.

Part IV
Planificació temporal

8 Planificació

8.1 Fases del projecte

El projecte es va dividir en tres fases (inicial, intermèdia i final), en les quals es van dur a terme les diferents parts que el componen. A la fase inicial, es va escriure un document que defineix el projecte, a la intermèdia es va implementar el projecte tal com es va definir al document i a la fase final es van recollir i presentar les conclusions. Cada fase es va dividir en una sèrie de tasques concretes amb una certa duració i uns certs requisits de forma que van permetre establir un ordre de realització sense conflictes.

8.1.1 Fase inicial: *Define*

A la fase inicial es va definir el projecte. Per fer-ho, es va començar fent recerca sobre el problema que es volia solucionar, les possibles solucions existents i l'estat de l'art relacionat. Un cop recollit aquest material, es va proposar l'abast i els objectius de la solució que es volia implementar. Tal com s'ha indicat en el capítol anterior, es va seguir una estructura de *milestones* i tasques:

- ***Milestone* 1A: Abast i contextualització (19/09 - 26/09)**
  - **Investigar el context del problema i els actors implicats** (3h): A quin camp pertany el problema? Qui se'n veuria beneficiat? Per què és rellevant? Quin és l'estat en l'actualitat?
  - **Investigar l'estat de l'art relacionat** (15h): Existeixen solucions al problema? És convenient crear una solució nova o reutilitzar altres ja creades? En quin estat es troben els camps relacionats en l'actualitat? Quin són els avenços recents?
  - **Anàlisi del problema i la solució** (2h): Per què és un problema que s'ha de solucionar? Definir en termes concrets el problema, separant les parts que siguin independents. Definir la solució, el seu abast i els objectius que es volen complir.
  - **Definir metodologia** (2h): Investigar metodologies de treball i eines que permetin dur-les a terme. Decidir quina metodologia s'utilitzarà per a implementar la solució.
- ***Milestone* 2A: Planificació (26/09 - 3/10)**
  - **Definició de les tasques a realitzar, alternatives i plans d'acció** (5h): Llistat i definició de les tasques i *milestones*. Consideració de possibles desviacions i plans d'acció alternatius.

- **Recursos i requeriments** (1h): Definir de quins recursos farà falta disposar o quins requisits farà falta complir per a poder realitzar les tasques.
- **Diagrama de *Gantt*** (2h): Crear el diagrama de *Gantt* de les tasques definides, juntament amb la seva estructura i requeriments.

- ***Milestone* 3A: Gestió econòmica i sostenibilitat (3/10 - 9/10)**
  - **Investigar cost dels recursos necessaris** (2h): Preu del hardware i software, salaris i altres despeses.
  - **Calcul del cost total del projecte** (2h): Utilitzar el cost dels recursos per calcular el cost final del projecte.
  - **Viabilitat i sostenibilitat** (4h): Considerar si el projecte és viable. Estudi de sostenibilitat en els diversos aspectes (econòmic, social i ambiental). Matriu de sostenibilitat.

- ***Milestone* 4A: Presentació preliminar (9/10 - 16/10)**
  - **Preparació del contingut de la presentació** (2h)
  - **Gravació de la presentació** (2h): Fer diversos assajos i comprovar que es compleixen tots els requisits.

- ***Milestone* 5A: Presentació oral i document final (12/09 - 23/10)**
  - **Document final** (15h): Utilitzar el *feedback* de les entregues parcials per sintetitzar el document final.
  - **Presentació oral** (3h): Crear material per a una presentació oral.

8.1.2 Fase intermèdia: *Establish*

En aquesta fase s'investiguen i s'implementen els mecanismes necessaris per a solucionar el problema, fent ús de la metodologia, les tasques i les eines especificades a la fase inicial:

- ***Milestone* 1B: Exploració i anàlisi de dades (19/09 - 23/10)**
  - **Instal·lació i proves de les llibreries a utilitzar** (6h): Llegir la documentació, descarregar i instal·lar les llibreries, executar proves, entendre les opcions, l'entrada i la sortida.
  - **Crear eines temporals d'exploració** (12h): Scripts i programes destinats a visualitzar o explorar les dades.
  - **Investigar algorismes i estructures de dades** (12h): Quins algorismes podem fer servir per analitzar les dades? Quines estructures necessitem per fer servir aquests algorismes?
  - **Proves de concepte** (16h): Implementacions bàsiques a manera d'exemple per comprovar que es poden obtenir resultats satisfactoris a partir dels mecanismes escollits.

- ***Milestone* 2B: Machine Learning (23/10 - 30/10)**
  - **Escollir i recopilar dades d'entrenament** (4h): Fer servir els mitjans necessaris per cercar i recopilar la informació necessària per a l'entrenament, ja sigui descarregant arxius, creant scripts, etc.
  - **Generar el model** (4h): Fent servir els algoritmes escollits, generar un model entrenant un sistema d'aprenentatge automàtic.
- ***Milestone* 3B: Disseny (30/10 - 6/11)**
  - **Definir requisits del sistema** (2h): Programes i llibreries necessàries.
  - **Definir casos d'ús** (4h): Definir com s'ha d'utilitzar el sistema.
  - **Disseny del sistema** (12h): Definir totes les parts del sistema i les seves interaccions mitjançant descripcions i diagrames.

8.1.3 Fase final: *Execute*

Per acabar, es recopilen els resultats i es preparen el document final i una presentació oral per tal d'exposar les conclusions del projecte.

- ***Milestone* 1C: Implementació (6/11 - 6/12)**
  - **Implementació** (62h): Fer ús del disseny proposat per crear el sistema definitiu.
  - **Validació** (8h): Comprovar que el resultat obtingut compleix amb els objectius proposats satisfactòriament.
- ***Milestone* 2C: Extensió (Opcional) (6/12 - 16/12)**
  - **Avaluació de possibles millores** (10h): Analitzar els resultats i proposar formes de millorar-los.
  - **Disseny i implementació de millores (iterativament)** (10h): Dissenyar i implementar de forma iterativa modificacions que puguin millorar el resultat obtingut.
- ***Milestone* 3C: Conclusions i resultats (10/12 - 10/01)**
  - **Document final** (16h): Incloure decisions preses i altres aspectes tècnics sobre la implementació.
  - **Presentació oral** (8h): Crear material per a una presentació oral incloent decisions, resultats i conclusions.

8.1.4 Seguiment (SE)

Paral·lelament a la realització del projecte, es farà un seguiment setmanal amb el director.

8.2 Pla d'acció

8.2.1 Paral·lelisme de les tasques

Tot i que moltes de les tasques es podrien haver paral·lelitzat i fer-se alhora, el projecte havia de ser realitzar per una única persona. Per això, només s'hi van paral·lelitzar aquelles tasques que s'havien de fer alhora de forma estricta, establint unes dependències artificials a les altres tasques per definir un ordre de compliment.

8.2.2 Possibles desviacions

Existeixen una sèrie de possibilitats que podrien provocar canvis a les previsions de temps necessari per a completar les tasques. De donar-se algun d'aquests casos es poden prendre diferents decisions per a assegurar que el projecte finalitza dins del temps estipulat.

- **Problemes amb eines existents**: Aquest projecte és molt susceptible a aquest tipus de circumstància, ja que s'hi utilitzaran moltes eines ja existents diferents. Si es donés aquest cas, podem:
  - Preguntar a la comunitat i mentrestant, avançar la configuració d'una altra eina o posar la tasca en pausa i continuar amb un altre (ja que moltes tasques són paral·lelitzables).
  - Considerar una eina alternativa.
  - Dedicar més temps a aquesta tasca i reduir temps d'altres tasques d'expansió o complementaries.

- **Problemes amb els algorismes**: Existeix una gran component d'investigació i de prova i error. És difícil saber si un mètode funcionarà correctament per aquest problema abans de provar-lo. Per això, la tasca de trobar els algorismes i les estructures adients podria allargar-se fàcilment. En aquest cas podem:
  - Dedicar més temps a aquesta tasca i reduir temps d'altres tasques d'expansió o complementaries.
  - Seguir endavant amb el millor mètode aconseguit, encara que no sigui un resultat del tot acceptable. Intentar millorar un cop el projecte estigui més avançat per reduir el risc d'estancar-se.

8.3 Recursos humans

Per a realitzar aquestes tasques es necessitaran els següents recursos humans:

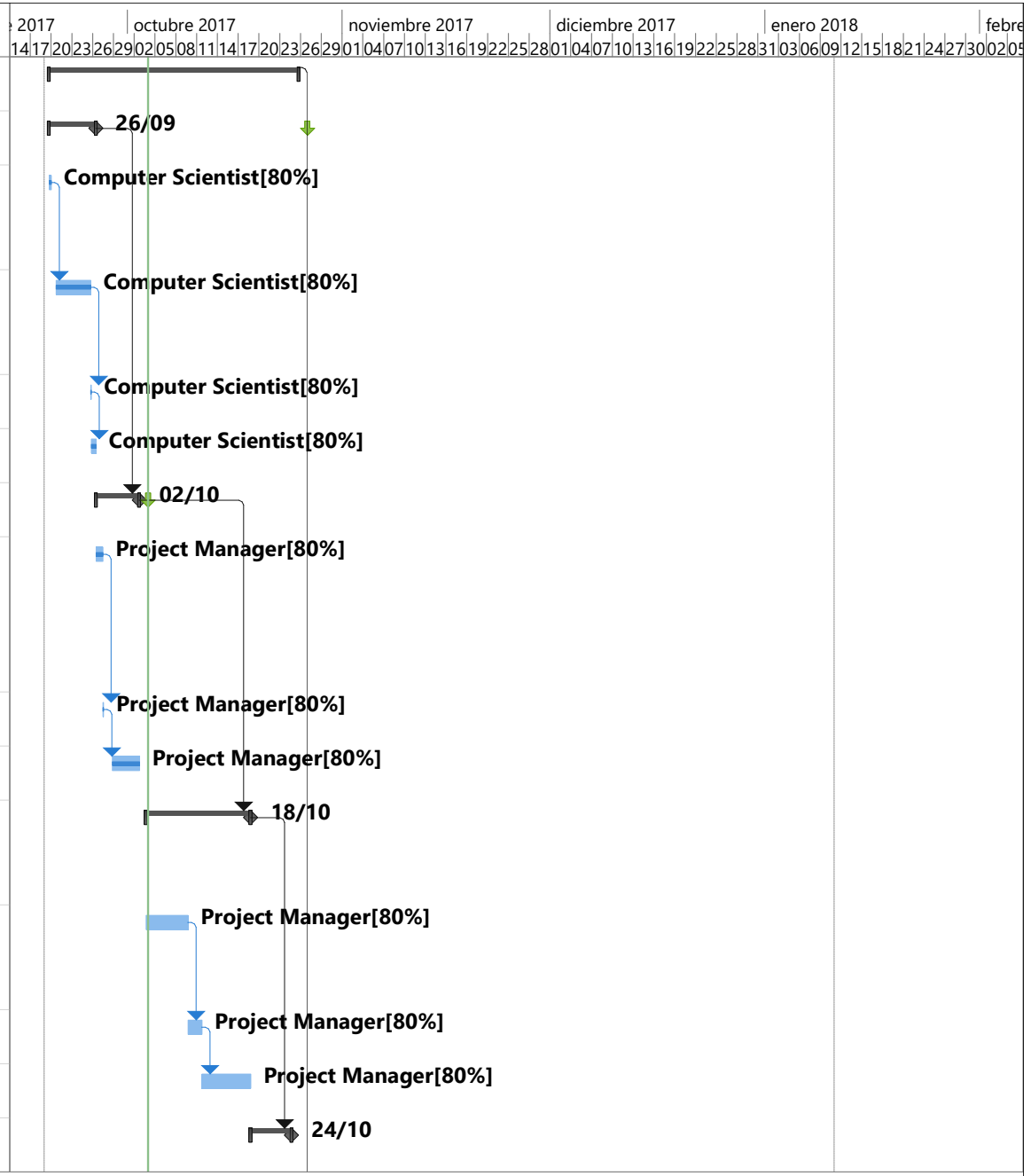
- ***Project manager***: Gestió del projecte i planificació.
- ***Computer scientist***: Persona amb coneixements sobre *Computer Science*.
- ***Software Engineer***: Persona amb capacitat de dissenyar i especificar sistemes de software.
- ***Programmer***: Persona amb capacitat d'implementar sistemes de software a partir d'una especificació.
- ***User***: Persona amb coneixements sobre les necessitats dels usuaris potencials finals.
- ***Director***: Ajudarà a guiar el projecte donant consell i avaluant la gestió i el desenvolupament.

Com que aquest projecte el realitza una única persona, ella mateixa haurà d'adoptar els diferents rols segons la tasca en la qual treballi, a excepció del rol de Director.

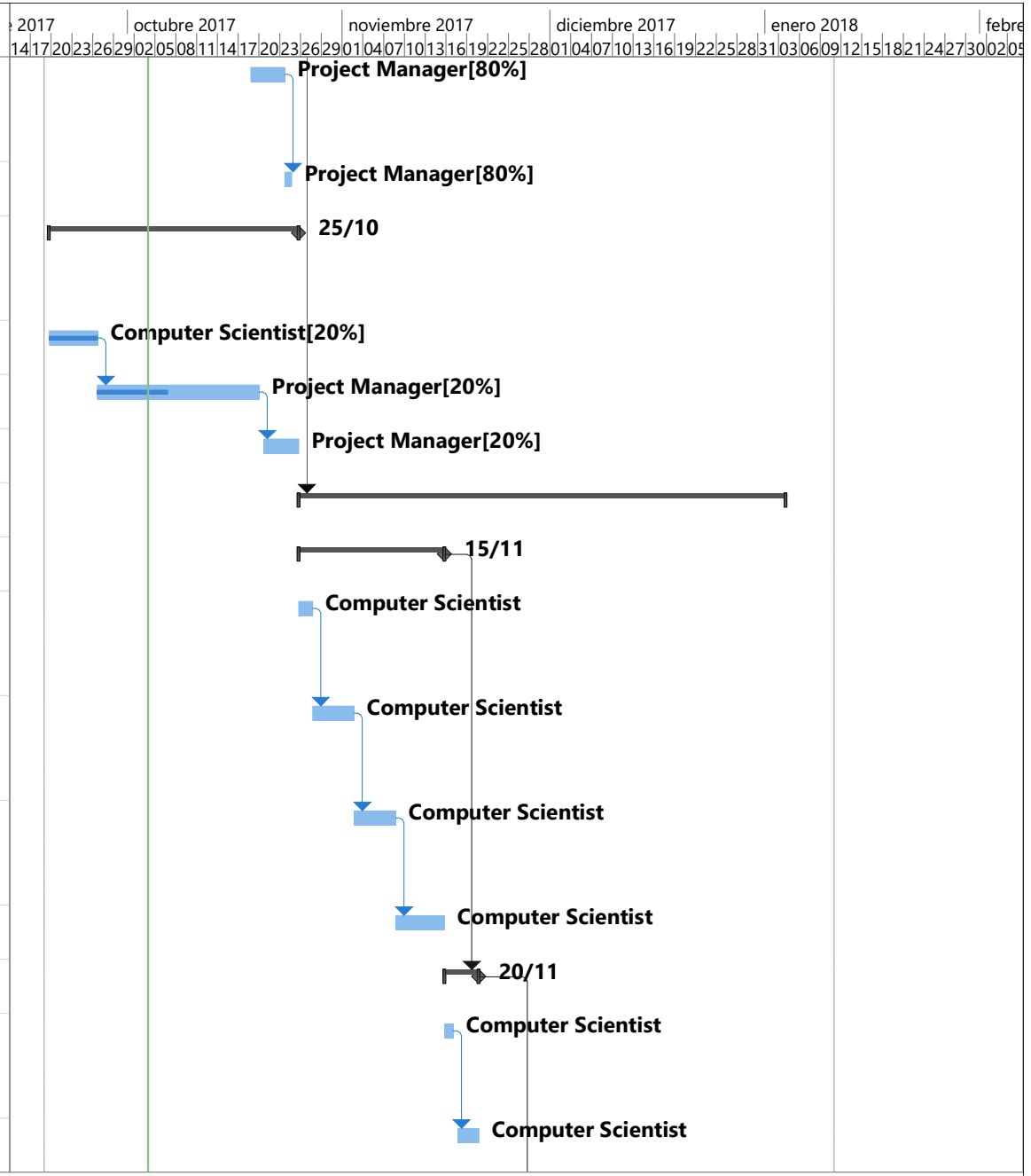
8.4 Diagrama de *Gantt*

El següent diagrama s'ha realitzat amb l'eina **Microsoft Project Professional 2016**. Els colors indiquen diferents nivells de risc. De menys a més risc: blau, taronja i vermell.

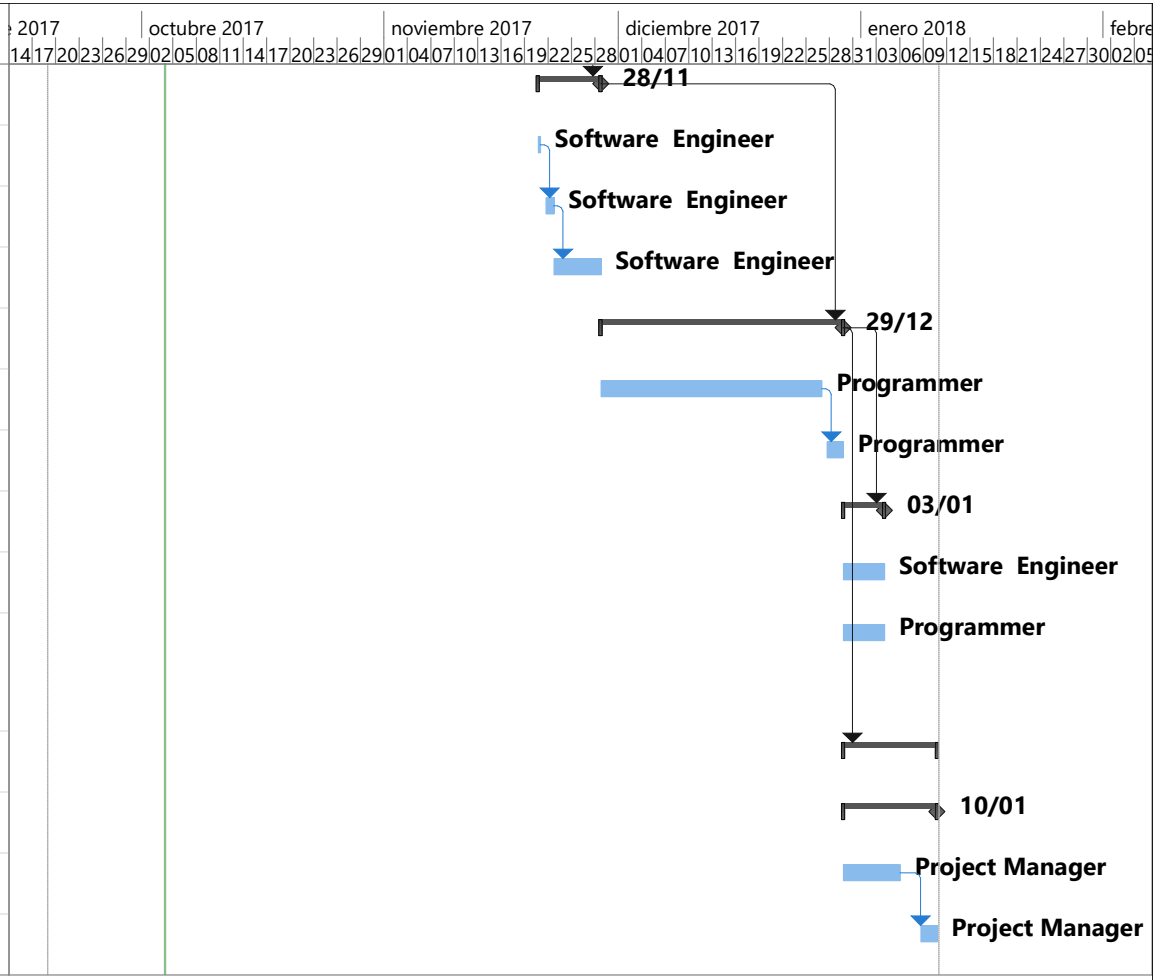
Id	Nombre de tarea	Trabajo	Duración
1	**Fase inicial**	**58,6 horas**	**23,63 días?**
2	**1A: Abast i contextualització**	**13,6 horas**	**4,5 días?**
3	Investigar el context del problema i els actors implicats	3 horas	0,47 días?
4	Investigar l'estat de l'art relacionat	6,6 horas	3 días
5	Anàlisi del problema i la solució	2 horas	0,31 días
6	Definir metodologia	2 horas	0,31 días
7	**2A: Planificació**	**16 horas**	**5 días**
8	Definició de les tasques a realitzar, alternatives i plans d'acció	5 horas	1,03 días
9	Recursos i requeriments	1 hora	0,16 días
10	Diagrama de Gantt	10 horas	1,84 días
11	**3A: Gestió econòmica i sostenibilitat**	**8,6 horas**	**9,78 días**
12	Investigar cost dels recursos necessaris	2,2 horas	1,66 días
13	Càlcul del cost total del projecte	2 horas	0,75 días
14	Viabilitat i sostenibilitat	4,4 horas	1,94 días
15	**4A: Presentació preliminar**	**2,4 horas**	**3,81 días**



Id	Nombre de tarea	Trabajo	Duración
16	Preparació del contingut de la presentació	0,4 horas	1 día
17	Gravació de la presentació	2 horas	0,31 días
18	**5A: Presentació oral i document final**	**18 horas**	**23,63 días**
19	Document final: CS	6 horas	5 días
20	Document final: PM	10 horas	9,13 días
21	Presentació oral	2 horas	1,25 días
22	**Fase intermèdia**	**162 horas**	**45,25 días**
23	**1B: Exploració i anàlisi de dades**	**46 horas**	**13,63 días**
24	Instalació i proves de les llibreries a utilitzar	6 horas	0,75 días
25	Crear eines temporals d'exploració	12 horas	1,5 días
26	Investigar algorismes i estructures de dades	12 horas	1,5 días
27	Proves de concepte	16 horas	2 días
28	**2B: Machine Learning**	**8 horas**	**3,38 días**
29	Escollir i recopilar dades d'entrenament	4 horas	0,5 días
30	Generar el model	4 horas	0,5 días



Id	Nombre de tarea	Trabajo	Duración
31	**3B: Disseny**	**18 horas**	**5 días**
32	Definir requisits del sistema	2 horas	0,25 días
33	Definir casos d'ús	4 horas	0,5 días
34	Disseny del sistema	12 horas	1,5 días
35	**4B: Implementació**	**70 horas**	**19,63 días**
36	Implementació	62 horas	7,75 días
37	Validació	8 horas	1 día
38	**5B: Extensió (Opcional)**	**20 horas**	**3,63 días**
39	Avaluació de possibles millores	10 horas	1,25 días
40	Disseny i implementació de millores	10 horas	1,25 días
41	**Fase final**	**24 horas**	**8 días**
42	**1C: Conclusions i resultats**	**24 horas**	**8 días**
43	Document final	16 horas	2 días
44	Presentació oral	8 horas	1 día



8.3 Recursos humans

Per a realitzar aquestes tasques es necessitaran els següents recursos humans:

- ***Project manager***: Gestió del projecte i planificació.
- ***Computer scientist***: Persona amb coneixements sobre *Computer Science*.
- ***Software Engineer***: Persona amb capacitat de dissenyar i especificar sistemes de software.
- ***Programmer***: Persona amb capacitat d'implementar sistemes de software a partir d'una especificació.
- ***User***: Persona amb coneixements sobre les necessitats dels usuaris potencials finals.
- ***Director***: Ajudarà a guiar el projecte donant consell i avaluant la gestió i el desenvolupament.

Com que aquest projecte el realitza una única persona, ella mateixa haurà d'adoptar els diferents rols segons la tasca en la qual treballi, a excepció del rol de Director.

8.4 Diagrama de *Gantt*

El següent diagrama s'ha realitzat amb l'eina **Microsoft Project Professional 2016**. Els colors indiquen diferents nivells de risc. De menys a més risc: blau, taronja i vermell.

8.5 Matriu d'assignació de responsabilitats

S'ha construït una matriu per descriure les responsabilitats de cada rol a les diferents parts del projecte.

Les lletres següents representen el nivell de responsabilitat:

- **R**: Responsable. Ha de gestionar la tasca i garantir la seva realització dins dels paràmetres establerts.
- **E**: Executor. Ha de realitzar la tasca tal com ha estat definida. Aporta les habilitats i l'esforç.
- **S**: Suport. Ajuda a la realització de la tasca d'alguna forma, sigui activament o mitjançant consell.

	Project Manager	Computer Scientist	Software Engineer	Programmer	User	Director
1A		R,E				S
2A	R,E					
3A	R,E					
4A	R,E					
1B		R,E				S
2B		R,E				S
3B			R,E		S	S
1C				R,E	S	S
2C			R	E	S	S
3C	R,E					S
SE	E					R,E

Taula. 1: Responsabilitat de cada rol per a cada *milestone*

Part V
Gestió econòmica i sostenibilitat

9 Identificació dels costos

En aquesta secció es detallen els costos que tenen els recursos necessaris per dur a terme el projecte descrit a les entregues anteriors.

9.1 Costos Directes

9.1.1 Recursos humans

Aquestes són les estimacions dels salaris de cada un dels rols involucrats en el desenvolupament, descrits a l'entrega anterior. S'han omès els costos dels rols d'usuari i director pel seu paper de suport. Els costos s'han estimat fent servir la pàgina *PayScale*.

Rol	**Salari**
Project Manager (PM)	38,62€
Software Engineer (SE)	28,73€
Computer Scientist (CS)	25,30€
Programmer (Pr)	17€

Taula. 2: Cost dels recursos humans

Fase inicial: Define	**77,07 horas**	**23,63 días?**	**2.714,30 €**	
1A: Abast i contextualització	**13,6 horas**	**4,5 días?**	**344,08 €**	
Investigar el context del problema i els actors implicats	3 horas	0,47 días?	75,90 €	Computer Scientist[80%]
Investigar l'estat de l'art relacionat	6,6 horas	3 días	166,98 €	Computer Scientist[80%]
Anàlisi del problema i la solució	2 horas	0,31 días	50,60 €	Computer Scientist[80%]
Definir metodologia	2 horas	0,31 días	50,60 €	Computer Scientist[80%]
2A: Planificació	**18,02 horas**	**5 días**	**695,73 €**	
Definició de les tasques a realitzar, alternatives i plans d'acció	5 horas	1,03 días	193,00 €	Project Manager[80%]
Recursos i requeriments	3,02 horas	1,66 días	116,73 €	Project Manager[80%]
Diagrama de Gantt	10 horas	1,84 días	386,00 €	Project Manager[80%]

3A: Gestió econòmica i sostenibilitat	**25,04 horas**	**9,79 días**	**966,54 €**	
Investigar cost dels recursos necessaris	10,62 horas	1,66 días	410,09 €	Project Manager[80%]
Càlcul del cost total del projecte	2 horas	0,75 días	77,20 €	Project Manager[80%]
Viabilitat i sostenibilitat	12,42 horas	1,94 días	479,26 €	Project Manager[80%]
4A: Presentació preliminar	**2,4 horas**	**3,81 días**	**92,64 €**	
Preparació del contingut de la presentació	0,4 horas	1 día	15,44 €	Project Manager[80%]
Gravació de la presentació	2 horas	0,31 días	77,20 €	Project Manager[80%]
5A: Presentació oral i document final	**18,01 horas**	**23,63 días**	**615,31 €**	
Document final: CS	6 horas	5 días	151,80 €	Computer Scientist[20%]
Document final: PM	10,01 horas	9,13 días	386,31 €	Project Manager[20%]
Presentació oral	2 horas	1,25 días	77,20 €	Project Manager[20%]

Fase intermèdia: Establish	**78 horas**	**22 días**	**1.883,34 €**	
1B: Exploració i anàlisi de dades	**46 horas**	**13,63 días**	**1.163,80 €**	
Instalació i proves de les llibreries a utilitzar	6 horas	0,75 días	151,80 €	Computer Scientist
Crear eines temporals d'exploració	12 horas	1,5 días	303,60 €	Computer Scientist
Investigar algorismes i estructures de dades	12 horas	1,5 días	303,60 €	Computer Scientist
Proves de concepte	16 horas	2 días	404,80 €	Computer Scientist
2B: Machine Learning	**8 horas**	**3,38 días**	**202,40 €**	
Escollir i recopilar dades d'entrenament	4 horas	0,5 días	101,20 €	Computer Scientist
Generar el model	4 horas	0,5 días	101,20 €	Computer Scientist
3B: Disseny	**24 horas**	**5 días**	**517,14 €**	
Definir requisits del sistema	4 horas	0,26 días	57,46 €	Software Engineer;User
Definir casos d'ús	8 horas	0,63 días	114,92 €	Software Engineer;User
Disseny del sistema	12 horas	1,5 días	344,76 €	Software Engineer

Fase final: Execute	**105 horas**	**27 días**	**2.362,05 €**	
1C: Implementació	**66 horas**	**19 días**	**1.122,00 €**	
Implementació	62 horas	7,75 días	1.054,00 €	Programmer
Validació	4 horas	0,5 días	68,00 €	Programmer
2C: Extensió (Opcional)	**15 horas**	**3,63 días**	**313,65 €**	
Avaluació de possibles millores	5 horas	0,63 días	143,65 €	Software Engineer
Disseny i implementació de millores	10 horas	1,25 días	170,00 €	Programmer
3C: Conclusions i resultats	**24 horas**	**8 días**	**926,40 €**	
Document final	16 horas	2 días	617,60 €	Project Manager
Presentació oral	8 horas	1 día	308,80 €	Project Manager
Reunions	**12 horas**	**49,75 días**	**463,20 €**	

Total recursos humans

$$2714,30€ + 1883,34€ + 2362,05€ + 463,20€ = \underline{7422,89€}$$

9.1.2 *Hardware*

Per realitzar aquest projecte s'utilitzaran una sèrie d'equips personals adquirits prèviament. A causa de l'ús que se li donarà, es consumirà part del seu temps de vida. Per tant afegirem al cost total el cost del temps de vida que es consumirà realitzant el projecte. Considerem el *hardware* com a cost directe perquè s'utilitzarà aquest equip específicament per a la realització del projecte, i a cap altra tasca durant la jornada laboral. Per calcular l'amortització, calculem el cost per hora del hardware i ho multipliquem pel número d'hores que es dedicaran al projecte:

$$\frac{Preu}{anys\_utils * 251 * 8} * hores\_projecte * unitats$$

Hi ha 251 dies laborals en un any i considerem jornades de 8 hores. En total es treballaran 266 hores. En aquest cas, les unitats és la fracció del temps en la que serà utilitzat.

Hardware	**Preu**	**Unitats**	**Vida útil**	**Amortització**
PC	1200,00€	70%	5 anys	22,25€
Portàtil	500,00€	30%	5 anys	3,97€
Perifèrics	200,00€	1	5 anys	5,3€
Total				**31,52€**

Taula. 3: Total Costos Hardware

9.1.3 *Software*

Tot i que a la realització del projecte s'utilitza software comercial, s'utilitzen llicències proporcionades per la UPC i, per tant, no repercuteixen en el cost del projecte.

9.1.4 Total Costos Directes

	Import
Recursos Humans	7422,89€
Hardware	29,08€
Software	0€
TOTAL	**7451,97€**

Taula. 4: Total Costos Directes

9.2 Costos Indirectes

9.2.1 Consum Elèctric

L'equipament i les instal·lacions fetes servir per a realitzar el projecte tenen un consum elèctric. Cal especificar que l'ordinador i els perifèrics consumeixen al voltant de 300W. Amb un preu de 0.133€/kWh, l'energia elèctrica consumida a la realització del projecte tindrà un cost de

$$0.133€/\text{kWh} * 0.3\text{kW} * 266\text{h} = \underline{10,61€}$$

9.2.2 Quota Internet

L'ús d'Internet és essencial en aquest projecte, ja sigui per descarregar eines com per fer recerca d'informació i dades. En aquest cas la quota d'Internet és de 30€al mes, i un mes té al voltant de 720 hores. Per tant el preu per hora és de 30/720 = 0.041. En total, el cost de l'Internet per realitzar el projecte és de

$$0.041€/\text{h} * 266\text{h} = \underline{10,91€}$$

9.2.3 Total Costos Indirectes

	Import
Consum elèctric	10,61€
Quota Internet	10,91€
TOTAL	**21,52€**

Taula. 5: Total Costos Indirectes

9.3 Cost dels riscos

A l'entrega anterior es van detallar els possibles riscos que es preveien per a aquest projecte. En cas de passar alguna d'aquestes complicacions, provocaria que s'haguessin d'usar més recursos dels previstos. Per això, afegim al cost del projecte un marge de risc de forma preventiva.

	Import
Problemes amb les eines	10%
Problemes amb els algorismes	15%
TOTAL	**25%**

Taula. 6: Total Costos Riscos

9.4 Cost total del projecte

Per últim, calculem el cost total del projecte fent ús de tots els costos identificats, afegint una contingència i el cost dels riscos.

	Import
Total Cost Directe	7451,97€
Total Cost Indirecte	21,52€
CD+CI	7473,49€
Contingència (15%)	1121,02€
CD+CI+Contingència	8647,98€
Total Riscos	2162€
TOTAL	**10809,98€**

10 Control de gestió

Com que hem previst que el projecte podria desviar-se de la planificació establerta si succeïssin algun dels riscos descrits, cal determinar uns mecanismes per tal de calcular el desviament. Amb aquest desviament podem saber si la gestió del projecte està sent adequada i valorar possibles modificacions o preveure situacions en projectes futurs.

Per calcular aquest desviament podem calcular la diferència entre la quantitat d'hores que s'han trigat a concloure una tasca i la quantitat que havia estat planejada. D'aquesta forma també podem calcular la desviació del pressupost, calculant el cost real i l'estimat.

Per calcular el desviament en un moment donat farem un informe indicant el següent:

- **Càlcul desviacions:**
  - Desviament de mà d'obra en cost = (cost estimat - cost real) * hores reals
  - Desviament de mà d'obra en consum = (consum d'hores estimat - consum d'hores reals) * cost real
  - Desviament total en mà d'obra = cost total estimat de mà d'obra - cost total real de mà d'obra
  - Desviament total de cost = cost total - cost real
- **Avaluació:**
  - S'ha produït alguna desviació rellevant?
  - **On** s'ha produït la desviació?
  - **Per què** s'ha produït la desviació?
  - **Quanta** desviació hi ha?

11 Anàlisi de sostenibilitat

En aquest apartat s'estudia la sostenibilitat del projecte. Per fer-ho, s'analitzarà l'efecte que té a l'economia, a la societat i al medi ambient contestant una sèrie de preguntes i s'hi calcularà una puntuació orientativa. Finalment es calcularà la matriu de sostenibilitat per obtenir una puntuació general de la sostenibilitat del projecte.

11.1 Econòmica

Com que aquest és un projecte de concepte més que un producte, no s'ha realitzat un estudi de mercat per tal d'avaluar els possibles usuaris, les seves despeses, les seves necessitats, etc. Per aquesta raó no podem justificar si el programa seria viable econòmicament o no. Així i tot, si assumim que l'aplicació acaba obtenint bons resultats, probablement ho seria, ja que:

- Al ser un programa, l'escalabilitat és infinita. Podem distribuir tantes còpies o llicències com calgui sense cost afegit.
- Existeixen milers d'empreses i professionals dedicats a la producció de contingut 3D. Si l'aplicació funciona, fins i tot amb tarifes molt assequibles, es recuperaria el cost del projecte molt ràpidament.
- No existeix una competència clara. No hi ha altres aplicacions similars.

Pel que fa a l'estimació dels recursos i el temps necessari podria ser reduït si fos dut a terme per una persona amb més experiència en aquest camp, però probablement també seria més costós. A més, com que reutilitzem eines existents per a tots els processos on existeix alguna, sabem que els recursos són utilitzats de forma efectiva i no redundant.

S'ha de notar que existeixen riscos que podrien endarrerir el projecte i que no podem assegurar que el resultat sigui satisfactori, cosa que podria resultar en la pèrdua de la inversió.

11.2 Social

A Espanya existeixen moltes empreses de producció de CGI i videojocs. En concret, Barcelona, ciutat on es desenvolupa el projecte, és una de les localitzacions on més densitat d'aquest tipus d'empresa es pot trobar. La realització d'aquest projecte podria ajudar a les empreses més petites a ser més competitives, permetent la generació de contingut utilitzant menys recursos i efectivament reduint la desigualtat entre empreses. Existeix el perill, però, que altres empreses més grans vegin oportú prescindir de personal a causa de l'estalvi de recursos que podria generar. Cal notar que els llocs de feina perduts possiblement serien de caràcter més mecànic i de menys qualitat (per exemple interpoladors).

Un altre aspecte a considerar, és que si l'aplicació esdevé en un increment de la producció de contingut 3D, podria donar-se un escenari al qual les empreses prioritzen la quanti-

tat sobre la qualitat i els consumidors d'aquest tipus de cultura podrien veure's afectats negativament.

A part d'empreses privades, aquesta aplicació podria tenir un impacte en sectors públics, com ara museus o altres llocs turístics que podrien utilitzar assistents virtuals per oferir assistència o guia. D'aquesta forma la societat podria veure's beneficiada.

A més, amb la realització d'aquest projecte es pretén guanyar experiència i aprendre més sobre els camps de la intel·ligència artificial i la generació d'imatges per ordinador.

En tot cas, aquest projecte hauria de ser una eina que permeti als usuaris generar contingut que no podrien generar d'un altre forma, o ajudant a fer que dediquin el temps a les tasques més creatives i interessants.

11.3 Mediambiental

L'impacte mediambiental del projecte és negligible, ja que l'únic impacte esdevé de l'energia consumida per l'equip utilitzat per realitzar-lo o l'equip que executi l'aplicació un cop acabada. En els dos casos, els equips estarien realitzant moltes altres tasques al mateix temps, diluint la contribució d'aquesta aplicació en concret.

Potencialment, amb aquest programa es podria reduir el temps de crear productes audiovisuals. Com que la quantitat de productes que es realitzen és molt gran, es podria reduir en gran quantitat els recursos totals utilitzats (consum elèctric, amortització d'equip). Això provocaria un impacte positiu en el medi ambient. Probablement, però, el que passaria realment és que s'incrementi la producció en general de forma que amb els mateixos recursos es generi més contingut.

La vida útil del programa depèn dels avenços que es produeixin en els camps relacionats, que podrien provocar la seva obsolescència.

En resum, aquest projecte té potencial per provocar un impacte positiu al medi ambient, però depèn dels usuaris i del consum dels productes generats utilitzant aquesta eina.

11.4 Matriu de sostenibilitat

Tenint en compte l'anàlisi realitzat obtenim la següent matriu de sostenibilitat.

	PPP	**Vida Util**	**Riscos**
Ambiental	Consum: 9/10	Petjada ecològica 19/20	Riscos ambientals 0/-20
Econòmica	Factura 8/10	Plan de viabilitat 14/20	Riscos econòmics -10/-20
Social	Impacte personal 9/10	Impacte social 15/20	Riscos socials -6/-20
Rang	26/30	48/60	-16/-60
	58		

Taula. 7: Matriu de sostenibilitat

Part VI
Execució del projecte

12 Canvis i Dificultats

12.1 Canvis a la gestió

En començar a treballar en el projecte, es va veure que, tenint el diagrama de *Gantt*, el seguiment amb *Kanban* era redundant i tediós, ja que s'havien de crear les tasques, ordenar-les i intentar mantenir la consistència amb el diagrama de *Gantt*. Per aquesta raó, es va continuar només amb la planificació gestionada al *Microsoft Project*.

12.2 Dificultats

12.2.1 Selecció d'esquelet d'interès

OpenPose retorna, per cada *frame*, una sèrie de punts que indiquen les posicions de cada part del cos de totes les persones a la imatge. Com que ho analitza per cada *frame* de forma independent, a un *frame* concret poden aparèixer o desaparèixer esquelets extres, o canviar l'ordre dels que hi havia al *frame* anterior. Tot i que el *dataset* utilitzat està format per vídeos on només apareix una persona, hi ha hagut situacions on es detectaven esquelets extra. Aquest problema s'ha solucionat calculant l'àrea dels *bounding box* dels esquelets que apareixen a un *frame* i guardant la informació de l'esquelet més gran. A part d'això, *OpenPose* assigna un *score* que representa la confiança que la posició d'aquella part és correcte. D'aquesta forma, en el càlcul del *bounding box* només es tenen en compte aquelles parts que superen un cert llindar de *score*. Això permet filtrar possibles esquelets grans que es puguin trobar al fons de la imatge per equivocació.

12.2.2 Associació de finestres de diferents *features*

Inicialment, en calcular les *features* sobre les postures dels esquelets, es feien servir finestres delimitades per un cert número de *frames*. Això, però, va convertir-se en un problema, ja que els vídeos d'entrenament tenien *framerates* diferents, provocant canvis a la duració de les finestres. La solució va consistir a utilitzar finestres definides per durada en temps (ms). Aquest canvi va introduir bastant complexitat, ja que per generar els fitxers d'entrenament, s'han d'associar els valors les finestres de diferent mida dels *features* de l'àudio amb els *features* de les postures que succeeixen al mateix moment.

12.2.3 Exploració inicial

Inicialment, es va fer un codi bàsic per fer proves de concepte. Aquest codi, però, va resultar limitant, ràpidament, i no es va poder arribar a realitzar el procés complet. Per

aquesta raó, es va decidir començar de nou amb una arquitectura més flexible en ment, suposant que al final es produirien resultats acceptables.

12.2.4 Problemes amb llibreries

Inicialment, la intenció era construir tant models de classificació com de regressió, però quan el projecte ja es trobava en un estat avançat, va haver-hi problemes per dur a terme l'entrenament de regressió. La llibreria de GRT donava un error en un cert punt de l'execució per a qualsevol cas. Es va decidir buscar una alternativa, però entre que ja era tard i la investigació de la nova opció requeria temps, va quedar fora de les possibilitats del projecte. Un altre problema va aparèixer a l'intentar modificar la mida de les finestres dels *functionals* a *openSMILE*. Després de fer vàries proves, es va descobrir que les finestres no responien a canvis continus de la mida, sinó que només canviaven cada 300 ms, aproximadament. A més, per alguna raó, en fer més gran la finestra es retornaven molts valors invàlids.

12.2.5 Problemes amb *Blender*

Al programar un *add-on* per *Blender*, existeixen tota una sèrie de factors que poden arribar a ser força problemàtics de no ser tinguts en consideració. S'ha de vigilar a quins *paths* busca *Blender* els mòduls, i quin és el *current working directory*, ja que els resultats poden ser diferents dels esperats. A més, intentar separar els mòduls de *Python* en *packages* (carpetes), sumat a la dificultat d'haver de registrar les noves classes de *Blender*, va resultar molt complicat. Al final, es va decidir no fer la separació per *packages* per motius pragmàtics.

12.3 Canvis a la planificació

La planificació va patir un canvi substancial, degut a un retràs important, a la meitat del projecte. Es va reestructurar l'ordre de les tasques i es va allargar el projecte. Com que a la planificació inicial es tenien en compte aquests riscos, el cost final del projecte no es veu afectat. Tot i això, com que el temps disponible es troba limitat per la data d'entrega i no es poden augmentar els recursos humans disponibles, es van haver de retallar els objectius finals.

12.4 Reestructuració de les tasques

	Fase inicial: Define	**77,07 horas**	**23,63 días?**	**2.714,30 €**	
	Fase intermèdia: Establish	**169 horas**	**32,13 días**	**4.378,60 €**	
✓	**1B: Exploració i anàlisi de dades**	**60 horas**	**18 días**	**1.518,00 €**	
✓	Instalació i proves de les llibreries a utilitzar	6 horas	0,75 días	151,80 €	Computer Scientist
✓	Crear eines temporals d'exploració	12 horas	1,5 días	303,60 €	Computer Scientist
✓	Investigar algorismes i estructures de dades	12 horas	1,5 días	303,60 €	Computer Scientist
✓	Proves de concepte	30 horas	3,75 días	759,00 €	Computer Scientist
	2B: Machine Learning	**79 horas**	**23,13 días**	**1.998,70 €**	
✓	Escollir i recopilar dades d'entrenament	4 horas	0,5 días	101,20 €	Computer Scientist
	Generar el model	75 horas	9,38 días	1.897,50 €	Computer Scientist
	3B: Disseny	**30 horas**	**9 días**	**861,90 €**	
✓	Definir requisits del sistema	2 horas	0,25 días	57,46 €	Software Engineer
✓	Definir casos d'ús	4 horas	0,5 días	114,92 €	Software Engineer
	Disseny del sistema	24 horas	3 días	689,52 €	Software Engineer
	Fase final: Execute	**161 horas**	**35,63 días**	**3.616,45 €**	
	1C: Implementació	**108 horas**	**32 días**	**1.836,00 €**	
	Implementació	100 horas	12,5 días	1.700,00 €	Programmer
	Validació	8 horas	1 día	136,00 €	Programmer
	2C: Extensió (Opcional)	**15 horas**	**3,63 días**	**313,65 €**	
	Avaluació de possibles millores	5 horas	0,63 días	143,65 €	Software Engineer
	Disseny i implementació de millores	10 horas	1,25 días	170,00 €	Programmer
	3C: Conclusions i resultats	**38 horas**	**12,38 días**	**1.466,80 €**	
	Document final	30 horas	3,75 días	1.158,00 €	Project Manager
	Presentació oral	8 horas	1 día	308,80 €	Project Manager
	Reunions	**12 horas**	**49,75 días**	**463,20 €**	

Figura. 14: Informació de les tasques després de la reestructuració

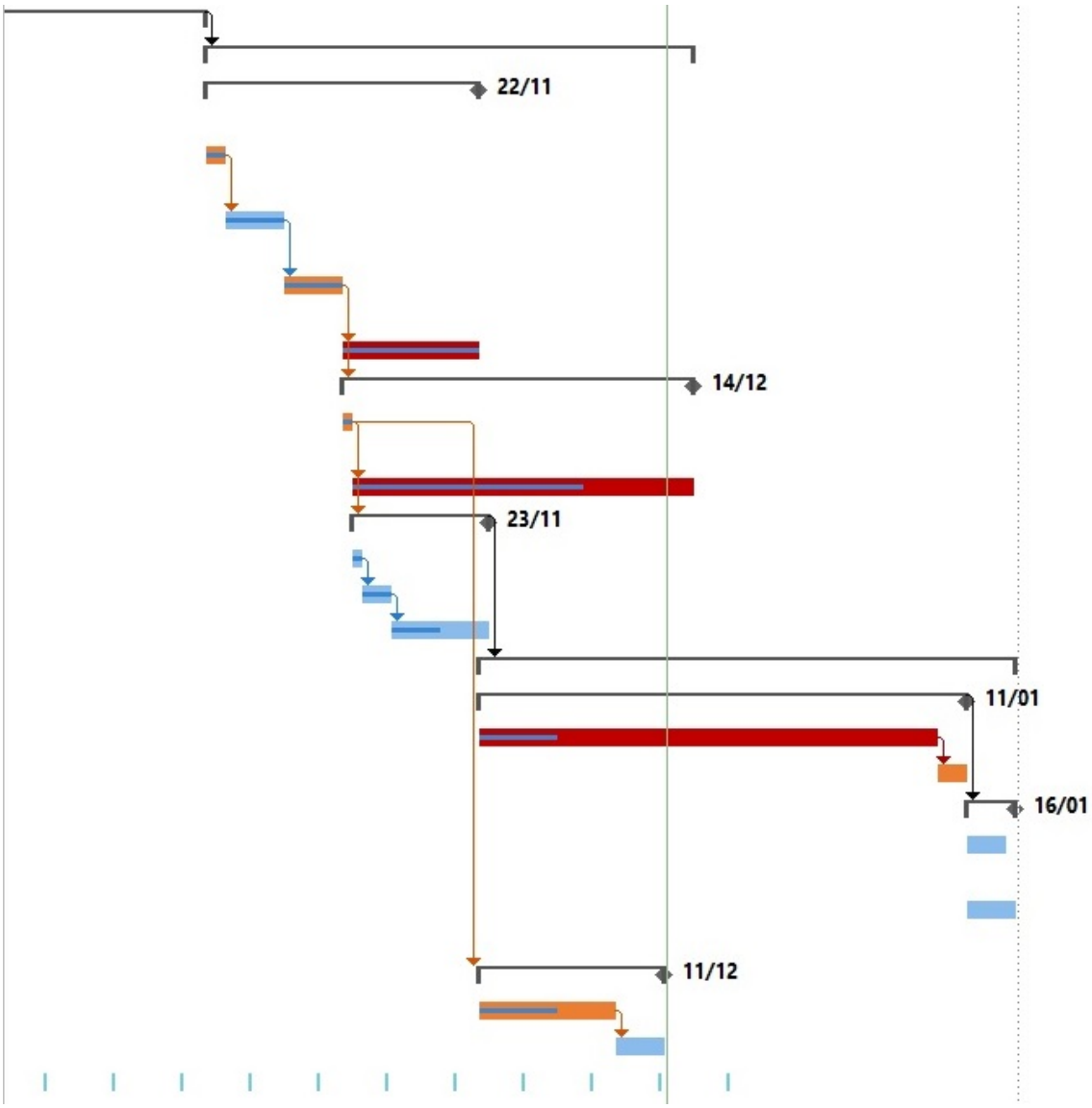


Figura. 15: Zona afectada del Diagrama de *Gantt* després de la reestructuració

Com podem veure a les imatges (14, 15), les proves de concepte van iniciar una sèrie de retards dins del projecte que van obligar a prendre mesures. Aquest retard va ser degut al fet que les proves de concepte creixien ràpidament i sense una escalabilitat correcta. Es va decidir començar el disseny abans d'hora al mateix temps que es feien les proves. Al final, però, es va decidir prescindir de les proves i començar a generar el model i implementar el programa al mateix temps, ja que era difícil preveure les necessitats del disseny i del model. D'aquesta forma es va començar una implementació iterativa d'estil àgil, a la qual es dissenya i implementa el programa d'acord amb les necessitats del moment. També es va aprofitar per anar documentant alguns dels problemes, algorismes i altres punts rellevants.

A la imatge (15), veiem l'estat en el qual es trobava el projecte el dia 12/12, moment en

el qual es va treballar en la reestructuració.

13 Desviació

13.1 Informe de desviació

- **Càlcul desviacions:**
  - Desviament de mà d'obra en cost = (cost estimat - cost real) * hores reals = 0 €
  - Desviament de mà d'obra en consum = (consum d'hores estimat - consum d'hores reals) * cost real

Rol	**Salari**	**D. Hores**	**D. Cost**
Project Manager (PM)	38,62€	0 h	0€
Software Engineer (SE)	28,73€	6 h	172,38€
Computer Scientist (CS)	25,30€	85 h	2150,30€
Programmer (Pr)	17€	42 h	714€
Total		133 h	**2322,88€**

  - Desviament total de cost = Desviament total en mà d'obra = cost total estimat de mà d'obra - cost total real de mà d'obra = 2322,88€
- **Avaluació:**
  - S'ha produït alguna desviació rellevant?
    Sí, el projecte ha crescut en una quantitat d'hores considerable.
  - **On** s'ha produït la desviació?
    Principalment, s'ha produït a les tasques 1B (*Proves de concepte*), 2B (*Generar el model*) i 1C (*Implementació*). Aquestes tasques ja havien estat considerades d'alt risc.
  - **Per què** s'ha produït la desviació?
    Hi han hagut problemes relacionats amb els riscos previstos a la planificació. Principalment problemes per entendre l'ús de les eines i correcció d'errors.
  - **Quanta** desviació hi ha?
    Hi ha un increment de 133 hores i 2322,88 €. Això representa un increment del 50% en hores i del 31% en cost (sense contingència ni riscos).

14 Cost de la desviació i cost final

La desviació té un cost total de 2322,88 €, però gràcies a la previsió de riscos i al marge de contingència, el projecte no ha sortit del pressupost (10809,98).

$$\text{Marge} = \text{Pressupost} - \text{Cost final} = 10809,98€ - (7473,49€ + 2322,88€) = 1013,61€$$

Part VII
Conclusió

15 Resultats

En aquest projecte hem aconseguit una eina fàcil d'utilitzar pels artistes i fàcil d'entrenar pels programadors o *data analysts*. En aquest apartat veurem els casos d'us principals, on es demostra la facilitat d'us del programa. Veurem com és el procés d'entrenament i d'animació. I a cada pas, veurem la informació (*output*) que ofereix el programa.

15.1 Procés d'entrenament

Per entrenar el model necessitem saber les *features* amb les que es vol entrenar, que això o podem obtenir amb el fitxer de configuració per defecte, o fer-ne una de pròpia. També es necessiten els vídeos d'entrenament.

Per modificar la configuració pròpia, cal afegir el següent: una forma de capturar la informació d'interès i una forma de classificar-la. Això ho aconseguim modificant els *windows values*, els *functionals* i els *labels groups*. Per exemple, suposant que volem aconseguir informació sobre el moviment del cap, hauríem d'afegir dins de *window values* una funció com la de la figura (16):

```python
def computeAngularVelocity(lastPartA, lastPartB, partA, partB):
    lastVector = lastPartB.pos - lastPartA.pos
    currVector = partB.pos - partA.pos
    #if we get a zero vector, we can't compute angular velocity
    if not lastVector.isZero() and not currVector.isZero():
        return lastVector.CwAngleWith(currVector)
    else:
        return 0

def headAngularVelocity(keypoints):
    windowSum = 0
    lastNeck = keypoints[0].pose.getPart(pose.BodyPartType.NECK)
    lastNose = keypoints[0].pose.getPart(pose.BodyPartType.NOSE)
    for keypoint in keypoints:
        neck = keypoint.pose.getPart(pose.BodyPartType.NECK)
        nose = keypoint.pose.getPart(pose.BodyPartType.NOSE)
        windowSum += computeAngularVelocity(lastNeck, lastNose, neck, nose)
        lastNeck = neck
        lastNose = nose

    #value for this window (mean of visited frames)
    windowValue = windowSum/float(len(keypoints))
    return windowValue


#valueName:function(keypoints->value)
WINDOW_VALUES = {
    'headAngularVelocity':headAngularVelocity
}

#funcName:function(window_values->value)
FUNCTIONALS = {
    'min':min,
    'max':max,
    'mean':statistics.mean,
    'median':statistics.median,
    'stdev':statistics.stdev,
}
```

Figura. 16: Funció *headAngularVelocity* extreta del fitxer de configuració per defecte

Les funcions que calculen *window values* reben un conjunt de *keypoints* (informació d'una postura a un *frame*) i han de retornar un valor. Per programar aquestes funcions podem fer ús del *package pose.py* que conté definicions de l'esquelet utilitzat per emmagatzemar la informació de les postures i el *package geometry.py*, que conté classes de punts i vectors. Les funcions que calculen *functionals* reben una llista de valors que han de resumir. Els mètodes del *package statistics* haurien de ser suficients per la majoria de casos. Aquests *functionals* s'aplicaràn per tots els *window values* definits. Per classificar les diferents seccions del vídeo, hem de definir grups de *labels*. Cada secció es classificarà en un label per cada grup. A efectes pràctics, es generarà un model de cadascun d'aquests grups (Figura 17).

Figura. 17: Contingut de *LABEL_GROUPS* del fitxer de configuració per defecte

D'aquesta forma podem classificar, per exemple, diferents parts del cos de forma independent. Per a cada *label* de cada grup, definim una expressió booleana que ens indica si aquella secció pertany a aquella classe. Aquestes expressions booleanes reben tant tots els *window values* (de la finestra avaluada) com els *functionals* (resum de tot el conjunt de finestres).

Un cop definit el nostre fitxer de configuració, disposem de eines per avaluar el seu comportament amb els vídeos d'entrenament. A la següents imatges (Figures 18,19) podem veure els valors dels *windows values* i els *functionals* per a un dels vídeos.

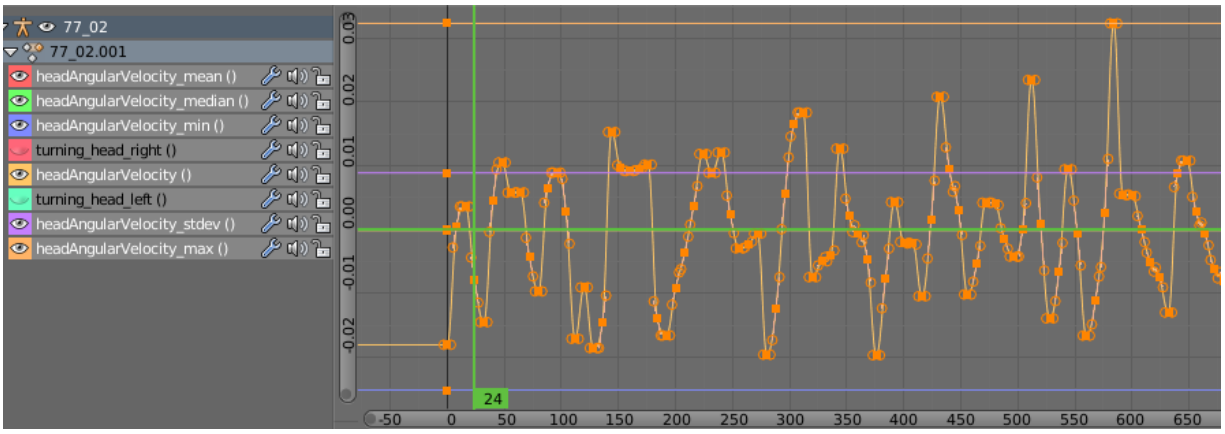


Figura. 18: Resultat de aplicar la configuració per defecte a un vídeo d'entrenament. Podem veure els diferents valors calculats.

Figura. 19: Classificació de les seccions del vídeo. Podem veure com en les seccions on hi han pics de *headAngularVelocity*, també s'ha assignat el *label* corresponent.

15.2 Procés d'animació

Un cop tenim un model generat, podem començar a animar avatars utilitzant discursos en format d'àudio. Com anteriorment, podem fer servir el fitxer de configuració per defecte o fer-ne un de propi. En aquest cas hem de definir com s'interpretarà la informació generada pel model. Hem de definir una funció (o més) rebent la informació dels *labels*, s'encarregui d'animar el avatar. En aquest cas podem fer ús de les funcions *getLabelIntervals* i *insertActions* definides al *package BlenderManager* per facilitar aquesta feina.

```python
def BaseAction(frames, armature):
    BlenderManager.insertAction('idle', 0, action_frame_end=len(frames))

def ActionAnimation(frames, armature):
    intervals = BlenderManager.getLabelIntervals('Head', '1', frames, threshold=2)
    for (start, end) in intervals:
        size = max(end-start, 50)
        BlenderManager.insertAction('talking', start, action_frame_end=size)


#animation passes
ANIMATIONS = [
    BaseAction,
    ActionAnimation
]
```

Figura. 20: A la imatge podem veure com s'inserixen animacions en relació als *labels* generats

D'aquesta forma, es genera una combinació de animacions que es corresponen amb l'àudio (Figura 21). Com podem veure a la imatge següent:

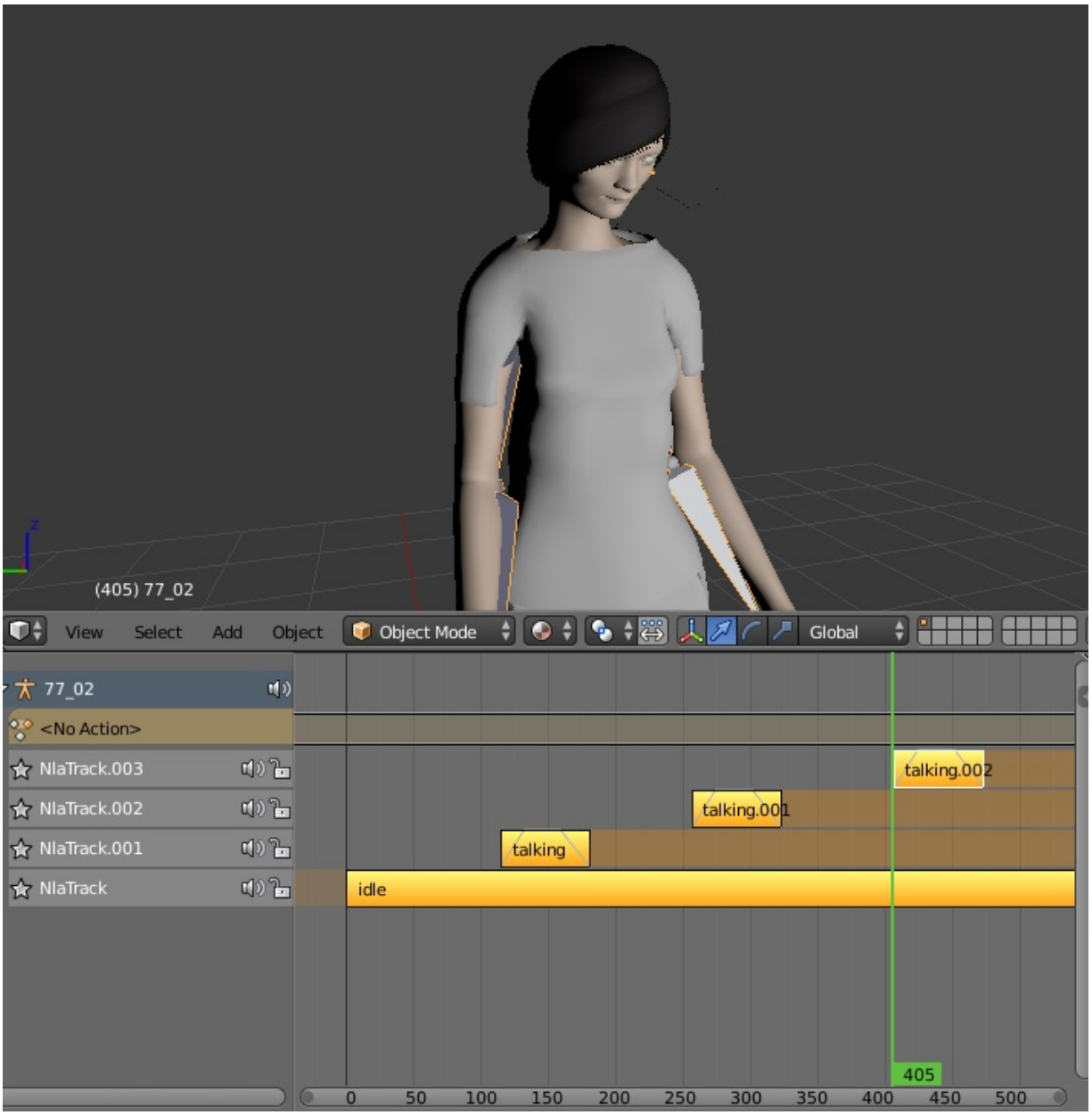


Figura. 21: A la imatge podem veure la composició d'animacions aconseguida

16 Conclusió

A mesura que ha anat avançant el projecte, els objectius han anat evolucionant a causa dels problemes i necessitats sorgits. En comptes d'intentar obtenir una solució única com s'havia proposat al principi, s'ha buscat aconseguir una plataforma que permeti arribar a diferents solucions de forma fàcil. Per tant, a l'hora d'avaluar l'èxit del projecte hauríem de reformular lleugerament els objectius.

16.1 Objectius finals

- **Aconseguir extreure informació dels vídeos d'entrenament en forma de postures humanes, animacions, velocitats, etc.**: Totalment realitzat. Per defecte s'extreu informació de velocitats de moviment, es calculen estadístics i s'assignen etiquetes per classificar els diferents esdeveniments. A més, tots aquests càlculs són fàcilment extensibles per l'usuari final.
- **Aconseguir crear un model que predigui la informació necessària per modificar posicions i velocitats per a algunes parts de l'esquelet a una animació existent**: Realitzat. Automàticament es generen diversos models que poden ajustar-se a diferents tipus de dades i es selecciona el millor d'ells. Aquests models aporten informació per poder modificar les animacions, però a l'haver fet servir únicament classificació, la informació que aporten és limitada i es delega molta feina a la interpretació d'aquesta.
- **Aconseguir que les modificacions generades donin una sensació versemblant que l'avatar està dient el mateix que a l'àudio d'entrada**: Parcialment realitzat. Sembla que les modificacions realitzades als avatars tenen algun tipus de relació amb l'àudio proporcionat, però no es pot dir que doni la sensació que reforci el missatge.
- **Crear una plataforma que pugui ajustar-se a les diferents necessitats dels usuaris i generar solucions diferents**: Aconseguit. La plataforma exposa uns fitxers de configuració que permeten definir fàcilment noves formes d'adquirir, processar i interpretar la informació.

17 Treball futur

Existeixen diversos punts del projecte que tenen marge per millorar o ser estesos:

- **Algorismes de classificació temporals**: Existeixen altres algorismes de classificació que tenen en compte l'ordre temporal de les mostres. En aquest cas, es possible que fer servir aquests tipus de algorismes ajudin a generar millors models.
- **Regressió**: Utilitzar models de regressió podria aportar variacions úniques a les animacions i ajudaria a disminuir la complexitat de les interpretacions actuals que es basen només en la classificació. El programa ara mateix ja genera fitxers d'entrenament per a regressió. Per tant només caldria afegir aquesta funcionalitat a l'executable que entrena els models. El model de regressió podria generar-se fent servir bé un meta-algorisme anomenat *Multidimensional Regression* que consisteix en produir múltiples models lineals o bé alguna mena de *Multi Layer Perceptron*, com ara *Deep Neural Nerworks* o *Convolutional Neural Nerworks*.
- **Reconeixement de la semàntica**: Alguns dels moviments realitzats a l'hora d'expressar-se van molt lligats a la semàntica. Es possible que fent servir tècniques

de reconeixement de la parla i la semàntica poguéssim obtenir informació rellevant per a produir animacions més versemblants.

- **Utilitzar transicions de diferents animacions**: Una altre forma efectiva de animar un personatge de forma procedural, seria controlant mescles d'animacions utilitzant diferents pesos i definides per a diferents parts del cos.

18 Valoració personal

Considero que el més important és que he après molts conceptes i eines que han despertat certa curiositat en mi i que m'agradaria aplicar en projectes personals en un futur pròxim. Si més no, opino que ha estat un projecte força difícil, ja que he hagut d'invertir la majoria del temps en investigar i aprendre. Tot i haver cursat les assignatures d'*Intel·ligència artificial* i Anàlisis de Dades i Explotació de la Informació, he trobat que no coneixia molts dels conceptes relacionats amb l'aprenentatge automàtic. També he hagut d'aprendre a fer servir moltes eines diferents (GRT, *OpenPose*, *openSMILE*) i possibles alternatives (*WEKA*, *scikit-learn*, *keras*) que no he acabat utilitzant, entre altres motius, per falta de temps.

Trobo que he aconseguit crear una plataforma capaç produir bons resultats, de molt fàcil ús i completa. A més, està feta de tal forma que, en cas de disposar de més temps, seria senzill realitzar els canvis necessaris que millorarien la qualitat substancialment (vegeu *Treball futur*).

Com a resum, he arribat a un conjunt de petites conclusions concretes:

- Un cop après com funciona el desenvolupament per *Blender* fent servir *Python*, és molt satisfactori i proporciona eines i opcions molt potents. A més, té molt bona documentació.
- M'he adonat que un projecte gran de *Python* pot desorganitzar-se ràpidament. Per futurs projectes hauré de tenir això en compte i investigar més sobre com organitzar el codi i aprendre *good practices*.
- Les tècniques de *Machine Learning* són molt útils i interessants, però només quan es comença a entendre com funcionen per darrere. Al principi del projecte he comès diversos errors de concepte que han fet que investigués més en elles.
- La llibreria GRT sembla molt interessant i m'agradaria investigar i provar en més profunditat les funcionalitats per la qual ha estat concebuda: la detecció de gesticulacions.
- Després d'haver investigat bastant sobre xarxes neuronals, al final no he pogut provar i aplicar-ho al projecte per falta de temps. És un tema que també m'agradaria estudiar en un futur.
- La gestió i planificació de projectes continua sent molt una tasca difícil tot i haver-ne fet ja uns quants. Cal recordar en el futur seguir aplicant planificacions i amb molt de marge de treball.

References

[1] J. Dunn, "Netflix subscribers over the years." [Online]. Available: http://www.businessinsider.com/netflix-subscribers-chart-2017-1

[2] L. Plunkett, "Nearly 40% Of All Steam Games Were Released In 2016," 2016. [Online]. Available: https://kotaku.com/nearly-40-of-all-steam-games-were-released-in-2016-1789535450

[3] Misix, "Special Effects Aren't Cheap: The Cost of CGI." [Online]. Available: https://misix.com/movie-quality-index-mar-7-mar-9-2014

[4] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent developments in openSMILE, the munich open-source multimedia feature extractor," in *Proceedings of the 21st ACM international conference on Multimedia - MM '13*. New York, New York, USA: ACM Press, 2013, pp. 835–838. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2502081.2502224

[5] T. Giannakopoulos, "pyAudioAnalysis: An Open-Source Python Library for Audio Signal Analysis," *PLOS ONE*, vol. 10, no. 12, p. e0144610, dec 2015. [Online]. Available: http://dx.plos.org/10.1371/journal.pone.0144610

[6] AAAC, "is17-compare — AAAC emotion-research.net - Association for the Advancement of Affective Computing." [Online]. Available: http://emotion-research.net/sigs/speech-sig/is17-compare

[7] H. Kaya and A. A. Karpov, "Introducing Weighted Kernel Classifiers for Handling Imbalanced Paralinguistic Corpora: Snoring, Addressee and Cold," in *Interspeech 2017*. ISCA: ISCA, aug 2017, pp. 3527–3531. [Online]. Available: http://www.isca-speech.org/archive/Interspeech{_}2017/abstracts/0653.html

[8] N. Takahashi, M. Gygli, and L. V. Gool, "AENet: Learning Deep Audio Features for Video Analysis." [Online]. Available: https://arxiv.org/pdf/1701.00599.pdf

[9] H. Kamper, A. Jansen, and S. Goldwater, "A Segmental Framework for Fully-Unsupervised Large-Vocabulary Speech Recognition," 2017. [Online]. Available: https://research.google.com/pubs/pub46041.html

[10] W. Williams, N. Prasad, D. Mrva, T. Ash, and T. Robinson, "Scaling recurrent neural network language models," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, apr 2015, pp. 5391–5395. [Online]. Available: http://ieeexplore.ieee.org/document/7179001/

[11] G. Saon, T. Sercu, S. Rennie, and H.-K. J. Kuo, "The IBM 2016 English Conversational Telephone Speech Recognition System," apr 2016. [Online]. Available: http://arxiv.org/abs/1604.08242

[12] G.-L. Chao, W. Chan, and I. Lane, "Speaker-Targeted Audio-Visual Models for Speech Recognition in Cocktail-Party Environments," 2016. [Online]. Available: https://pdfs.semanticscholar.org/ba06/93e465e39bbde6fc5832f8344817fb2da8fc.pdf

[13] J. R. Hershey, P. A. Olsen, S. J. Rennie, and A. Arron, "Audio Alchemy: Getting Computers to Understand Overlapping Speech - Scientific American." [Online]. Available: https://www.scientificamerican.com/article/speech-getting-computers-understand-overlapping/

[14] J. Shotton, A. Fitzgibbon, A. Blake, A. Kipman, M. Finocchio, B. Moore, and T. Sharp, "Real-Time Human Pose Recognition in Parts from a Single Depth Image," jun 2011. [Online]. Available: https://www.microsoft.com/en-us/research/publication/real-time-human-pose-recognition-in-parts-from-a-single-depth-image/

[15] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," in *CVPR*, 2017.

[16] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, "Hand Keypoint Detection in Single Images using Multiview Bootstrapping," in *CVPR*, 2017.

[17] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines," in *CVPR*, 2016.

[18] M. Everingham, "The PASCAL Visual Object Classes Homepage." [Online]. Available: http://host.robots.ox.ac.uk/pascal/VOC/

[19] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial Structures for Object Recognition." [Online]. Available: http://static.cs.brown.edu/people/pff/papers/blobrecJ.pdf

[20] P. Dollar, R. Appel, S. Belongie, and P. Perona, "Fast Feature Pyramids for Object Detection," *PAMI*, apr 2014. [Online]. Available: https://www.microsoft.com/en-us/research/publication/fast-feature-pyramids-for-object-detection/?from=http{%}3A{%}2F{%}2Fresearch.microsoft.com{%}2Fpubs{%}2F220572{%}2Fdollarpami14pyramids.pdf

[21] J. Carreira, P. Agrawal, K. Fragkiadaki, and J. Malik, "Human Pose Estimation With Iterative Error Feedback," pp. 4733–4742, 2016. [Online]. Available: https://www.cv-foundation.org/openaccess/content{_}cvpr{_}2016/html/Carreira{_}Human{_}Pose{_}Estimation{_}CVPR{_}2016{_}paper.html

[22] S. Suwajanakorn, S. M. Seitz, and I. Kemelmacher-Shlizerman, "Synthesizing Obama," *ACM Transactions on Graphics*, vol. 36, no. 4, pp. 1–13, jul 2017. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3072959.3073640

[23] S. Taylor, A. Kato, I. Matthews, and B. Milner, "Audio-to-Visual Speech Conversion using Deep Neural Networks." [Online]. Available: https://ueaeprints.uea.ac.uk/60483/1/Accepted{_}manuscript.pdf

[24] T. Karras, T. Aila, S. Laine, A. Herva, and J. Lehtinen, "Audio-driven facial animation by joint end-to-end learning of pose and emotion," *ACM Transactions on Graphics*, vol. 36, no. 4, pp. 1–12, jul 2017. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3072959.3073658

[25] K. Olszewski, J. J. Lim, S. Saito, and H. Li, "High-fidelity facial and speech animation for VR HMDs," *ACM Transactions on Graphics*, vol. 35, no. 6, pp. 1–14, nov 2016. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2980179.2980252

[26] Y. Cao, W. C. Tien, P. Faloutsos, and F. Pighin, "Expressive speech-driven facial animation," *ACM Transactions on Graphics*, vol. 24, no. 4, pp. 1283–1302, oct 2005. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1095878.1095881

[27] C.-C. Chiu and S. Marsella, "How to Train Your Avatar: A Data Driven Approach to Gesture Generation." Springer, Berlin, Heidelberg, 2011, pp. 127–140. [Online]. Available: http://link.springer.com/10.1007/978-3-642-23974-8{_}14

[28] N. Sadoughi and C. Busso, "Joint Learning of Speech-Driven Facial Motion with Bidirectional Long-Short Term Memory." Springer, Cham, aug 2017, pp. 389–402. [Online]. Available: http://link.springer.com/10.1007/978-3-319-67401-8{_}49

[29] C. Ding, L. Xie, and P. Zhu, "Head motion synthesis from speech using deep neural networks," *Multimedia Tools and Applications*, vol. 74, no. 22, pp. 9871–9888, nov 2015. [Online]. Available: http://link.springer.com/10.1007/s11042-014-2156-2

[30] J. Cassell, C. Pelachaud, N. I. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone, "ANIMATED CONVERSATION: Rule-Based Generation of Facial Expression, Gesture, Spoken Intonation for Multiple Conversational Agents," 1994. [Online]. Available: http://repository.upenn.edu/cis{_}reports

[31] N. Sadoughi and C. Busso, "Speech-Driven Animation with Meaningful Behaviors." [Online]. Available: https://arxiv.org/pdf/1708.01640.pdf

[32] A. W. Siegman and S. Feldstein, *Nonverbal Behavior and Communication.* Taylor and Francis, 2014.

[33] A. H. Anderson, E. G. Bard, C. Sotillo, A. Newlands, and G. Doherty-Sneddon, "Limited visual control of the intelligibility of speech in face-to-face dialogue," *Perception & Psychophysics*, vol. 59, no. 4, pp. 580–592, jun 1997. [Online]. Available: http://www.springerlink.com/index/10.3758/BF03211866

[34] M. Mori, "The Uncanny Valley," vol. 7, no. 4, pp. 33–35. [Online]. Available: http://www.movingimages.info/digitalmedia/wp-content/uploads/2010/06/MorUnc.pdf

[35] M. Sahidullah and G. Saha, "Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition," *Speech Communication*, vol. 54, pp. 543–565, 2012. [Online]. Available: http://www.cs.joensuu.fi/{~}sahid/Sahidullah{_}files/SPEECHCOM-2012.pdf

[36] J.-I. Biel, O. Aran, and D. Gatica-Perez, "You Are Known by How You Vlog: Personality Impressions and Nonverbal Behavior in YouTube." [Online]. Available: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2796/3220

[37] "Why Sitting Down Destroys You — Roger Frampton — TEDxLeamingtonSpa - YouTube." [Online]. Available: https://www.youtube.com/watch?v=jOJLx4Du3vU{&}feature=youtu.be

[38] "Intervención de Albiol PP ante el Parlamente de Cataluña 10 Octubre 2017 - YouTube." [Online]. Available: https://www.youtube.com/watch?v=Rs5LIxGUa6Y{&}feature=youtu.be